



---

UNIVERSIDAD DE VALLADOLID

TRABAJO FIN DE GRADO  
GRADO EN INGENIERÍA DE TECNOLOGÍAS DE TELECOMUNICACIÓN

# **Parametrización y clasificación de imágenes acústicas para biometría**

Autor:

**Andrés Martín Yeves**

Tutor:

**Alberto Izquierdo Fuente**

Valladolid, 20 de septiembre de 2022



---

<b>TÍTULO:</b>	<b>Parametrización y clasificación de imágenes acústicas para biometría</b>
<b>AUTOR:</b>	<b>D. Andrés Martín Yeves</b>
<b>TUTOR:</b>	<b>Dr. Alberto Izquierdo Fuente</b>
<b>DEPARTAMENTO:</b>	<b>Teoría de la Señal y Comunicaciones</b>

---

### **TRIBUNAL**

---

<b>PRESIDENTE:</b>	<b>Alberto Izquierdo Fuente</b>
<b>VOCAL:</b>	<b>Juan José Villacorta Calvo</b>
<b>SECRETARIO:</b>	<b>Ramón de la Rosa Steinz</b>
<b>SUPLENTE:</b>	<b>Alonso Alonso Alonso</b>
<b>SUPLENTE:</b>	<b>Jaime Gómez Gil</b>

### **Resumen del TFG:**

En biometría, las técnicas de clasificación se basan en la extracción de parámetros de un gran conjunto de datos. A partir de imágenes acústicas 3D obtenidas con un array de sensores MEMS de alta resolución, este trabajo pretende implementar un modelo paramétrico que asocie a cada imagen un conjunto de características. El objetivo es que estas permitan clasificar distintas personas en base a sus respuestas acústicas, utilizando un algoritmo SVM a través de LabVIEW.

Para ello, contar con un conjunto de imágenes de calidad para fines biométricos es imprescindible, por lo que el primer paso es considerar distintos aspectos y parámetros en relación con el sistema de adquisición y procesamiento de las imágenes. Seguidamente, se plantean distintos modelos basados en la respuesta espacial del array o en los máximos 3D presentes en las imágenes, así como distintas técnicas de extracción de parámetros que permitan obtener las características necesarias para entrenar los clasificadores. Finalmente, se evalúan los resultados conseguidos obteniendo para el mejor algoritmo una precisión de algo menos del 90%.

### **Palabras clave:**

Biometría acústica, array de micrófonos MEMS, modelo, clasificación, SVM, LabVIEW

## **Abstract:**

In biometrics, classification techniques are based on the extraction of parameters from a large dataset. Using 3D acoustic images obtained with a high-resolution MEMS sensor array, this work aims to implement a parametric model that associates a set of features to each image. The aim is to classify different people based on their acoustic responses using an SVM algorithm through LabVIEW.

For achieving this objective, it is essential to have a set of quality images for biometric purposes, so the first step is to consider different aspects and parameters in relation to the image acquisition and processing system. Subsequently, different models based on the spatial response of the array or on the 3D maxima present in the images are proposed, as well as different parameter extraction techniques to extract the necessary features to train the classifiers. Finally, the results obtained are evaluated, obtaining a precision slightly below of 90% for the best algorithm.

## **Keywords:**

Acoustic biometrics, MEMS microphone array, model, classification, SVM, LabVIEW

## **Agradecimientos:**

A mi familia y en especial a mi padre, por escuchar innumerables explicaciones nocturnas con cada nueva idea que surgía de madrugada.

A todo el equipo del GPA, Lara, Juanjo y en especial Alberto, por darme en todo momento la confianza, la formación y el apoyo necesario para sacar este proyecto adelante.

A mis amigos de siempre por la paciencia y por esos ratos de desconexión que a veces son tan necesarios para poder ver las cosas desde otro punto de vista.



---

**ÍNDICE DE CONTENIDO**


---

<b>1. Capítulo 1. INTRODUCCIÓN</b> .....	<b>1</b>
1.1 OBJETIVOS .....	2
1.1.1 Objetivo general .....	2
1.1.2 Objetivos específicos.....	2
1.2 ESTRUCTURA Y FASES DEL PROYECTO.....	3
1.2.1 Adquisición de imágenes acústicas .....	3
1.2.1 Diseño de un modelo paramétrico para biometría.....	3
1.2.2 Estimación de parámetros.....	3
1.2.3 Fase de entrenamiento y clasificación.....	4
1.2.4 Conclusiones y validación final.....	4
1.3 ESTADO DEL ARTE.....	4
<b>2. Capítulo 2. MARCO TEÓRICO</b> .....	<b>6</b>
2.1 IMÁGENES ACÚSTICAS MULTIFRECUENCIA .....	6
2.2 BIOMETRÍA ACÚSTICA Y ANTROPOMETRÍA .....	9
2.3 ESTIMACIÓN, OPTIMIZACIÓN Y HEURÍSTICA .....	10
2.4 MACHINE LEARNING. SVM.....	11
2.5 ENTORNO DE DESARROLLO LABVIEW .....	14
<b>3. Capítulo 3. ADQUISICIÓN DE IMÁGENES</b> .....	<b>17</b>
3.1 SISTEMA DE ADQUISICIÓN HARDWARE.....	17
3.1.1 Tweeter .....	17
3.1.2 Mems array.....	18
3.1.3 Sistema de procesado.....	20
3.2 CONFORMADO Y PREPROCESADO .....	21
3.2.1 ANCHO de pulso.....	22
3.2.2 Frecuencia de la señal acústica.....	23
3.3 CONSIDERACIONES DE ADQUISICIÓN .....	25
3.3.1 Escenario .....	25
3.3.2 Posición del individuo y Límites espaciales.....	26
3.3.3 Discusión sobre las reflexiones acústicas .....	29
<b>4. Capítulo 4. DEFINICIÓN DEL MODELO</b> .....	<b>32</b>
4.1 APROXIMACIÓN SUCESIVA. MODELO REAL .....	32
4.1.1 Síntesis 1D.....	33
4.1.2 Síntesis 2D.....	35

---

4.1.3	Síntesis 3D.....	37
4.2	MODELO COMPLEJO 3D .....	39
4.2.1	Necesidad de un modelo más riguroso.....	40
4.2.2	Presentación del modelo complejo.....	42
4.2.3	Análisis del nuevo modelo .....	44
4.3	MODELO INTELIGENTE DE MÁXIMOS 3D .....	45
4.3.1	Estudio de los blancos significativos de un individuo.....	45
4.3.2	Presentación del modelo inteligente.....	49
4.4	CONSIDERACIONES 4D .....	50
5.	<b>Capítulo 5. EXTRACCIÓN DE PARÁMETROS .....</b>	<b>51</b>
5.1	PLANTEAMIENTO INICIAL DE ESTIMACIÓN. LSE.....	52
5.2	OPTIMIZACIÓN. ALGORITMOS DE BÚSQUEDA .....	55
5.2.1	Búsqueda exhaustiva .....	56
5.2.2	Búsqueda aleatoria.....	57
5.3	SQP. OPTIMIZACIÓN NO LINEAL .....	59
5.4	ALGORITMO DE BÚSQUEDA INTELIGENTE.....	61
5.4.1	Búsqueda .....	61
5.4.2	Asociación. Contenedores adaptativos.....	62
6.	<b>Capítulo 6. IDENTIFICACIÓN. SVM.....</b>	<b>65</b>
6.1	ENTRENAMIENTO .....	67
6.1.1	Definición y extracción de características.....	67
6.1.2	Hiperparámetros .....	71
6.1.3	Validación.....	72
6.2	CLASIFICACIÓN .....	72
6.2.1	Métricas de Evaluación: .....	72
6.2.2	Exposición de Resultados.....	75
6.2.3	Análisis de resultados .....	78
7.	<b>Capítulo 7. CONCLUSIONES .....</b>	<b>80</b>
7.1	LÍNEAS FUTURAS .....	81
8.	<b>BIBLIOGRAFÍA .....</b>	<b>82</b>

## - ÍNDICE DE FIGURAS

FIGURA 1: REPRESENTACIONES 2D DEL DIAGRAMA DE RADIACIÓN DE UNA ANTENA .....	6
FIGURA 2: TÉCNICA PULSO-ECO APLICADA CON UN SOLO TRANSCÉPTOR .....	7
FIGURA 3: ESTRUCTURA DE UN CONFORMADOR [6] .....	8
FIGURA 4: SISTEMA DE COORDENADAS ESFÉRICAS .....	8
FIGURA 5: PROYECCIÓN SOBRE EL PLANO AZIMUT-ELEVACIÓN DE UN BLANCO PUNTUAL .....	9
FIGURA 6: PROBLEMA DE CLASIFICACIÓN BINARIA EN 2 DIMENSIONES A TRAVÉS DE SVM [25].....	12
FIGURA 7 ENTRENAMIENTO Y VALIDACIÓN DEL IDENTIFICADOR.[14] .....	13
FIGURA 8: TÉCNICA DE VALIDACIÓN CRUZADA SFCV CON $S = 4$ [26] .....	13
FIGURA 9: MATRIZ DE CONFUSIÓN PARA CLASIFICADORES MULTICLASE [25].....	14
FIGURA 10: PROBLEMAS PRINCIPALES EN EL APRENDIZAJE SUPERVISADO .....	14
FIGURA 11: PANEL FRONTAL Y DIAGRAMA DE BLOQUES [29].....	15
FIGURA 12: PALETA DE CONTROLES .....	16
FIGURA 13: DISPOSICIÓN DEFINITIVA DEL SISTEMA DE ADQUISICIÓN .....	18
FIGURA 14: PRIMERA DISPOSICIÓN DEL SISTEMA DE ADQUISICIÓN .....	18
FIGURA 15: DIFERENCIA DE CAMINOS ENTRE LAS ONDAS INCIDENTE Y REFLEJADA .....	18
FIGURA 16: MEMS ARRAY.....	19
FIGURA 17: RESPUESTA FRECUENCIAL DE LOS MICRÓFONOS EN EL RANGO 20Hz - 10KHz.....	20
FIGURA 18: AMPLIFICADOR EXTERNO .....	20
FIGURA 19 MYRIO CONECTADA AL ARRAY ACÚSTICO MEMS [30] .....	21
FIGURA 20: DIAGRAMA DE RADIACIÓN DE UN ARRAY LINEAL CON $f = 8 KHz$ .....	24
FIGURA 21: DIAGRAMA DE RADIACIÓN DE UN ARRAY LINEAL CON $f = 12 KHz$ .....	24
FIGURA 22: DIAGRAMA DE RADIACIÓN DE UN ARRAY LINEAL CON $f = 14 KHz$ .....	24
FIGURA 23: DIAGRAMA DE RADIACIÓN DE UN ARRAY LINEAL CON $f = 16 KHz$ .....	24
FIGURA 24: DIAGRAMA DE RADIACIÓN DE UN ARRAY LINEAL CON $f = 20 KHz$ .....	25
FIGURA 25: DIAGRAMAS DE RADIACIÓN DE UN ARRAY LINEAL PARA $\theta = 30^\circ$ A DISTINTAS FRECUENCIAS: (A) 13 KHz, (B) 14 KHz, (C) 15 KHz, (D) 16 KHz .....	25
FIGURA 26: DIAGRAMA DE RADIACIÓN DEL ARRAY A 14KHz.....	26
FIGURA 27: POSICIONES PARA ADQUISICIÓN PLANTEADAS EN ANTERIORES ESTUDIOS [6] .....	27
FIGURA 28: POSICIÓN DEL INDIVIDUO ESCOGIDA PARA LAS CAPTURAS .....	27
FIGURA 29: DISPOSICIÓN DEL INDIVIDUO FRENTE AL ARRAY ACÚSTICO .....	28
FIGURA 30: IMPEDANCIAS ACÚSTICAS DE ALGUNOS TEJIDOS HUMANOS [16].....	30
FIGURA 31: SEÑALES $F1\theta$ , $F2\theta$ , $F3\theta$ Y $PPersona\theta$ .....	35
FIGURA 32: (A) IMAGEN ACÚSTICA TOMADA DE UN MANIQUÍ, (B) IMAGEN ÓPTICA DEL MANIQUÍ CAPTURADO .....	35
FIGURA 33: IMAGEN 2D $PPersona\theta, \varphi$ GENERADA A PARTIR DE 7 BLANCOS SINTÉTICOS .....	37
FIGURA 34: ANCHO DE BANDA A 3DB DEL DIAGRAMA DE RADIACIÓN DEL ARRAY .....	37
FIGURA 35: PRIMERAS PROPUESTAS DE PULSOS UTILIZADOS PARA SINTETIZAR LAS IMÁGENES 3D.....	38
FIGURA 36: PULSO GAUSSIANO DE 1MS UTILIZADO PARA SINTETIZAR LAS IMÁGENES 3D .....	38
FIGURA 37: PROYECCIONES DE UNA IMAGEN 3D $PPersona\theta, \varphi, r$ GENERADA A PARTIR DE 7 BLANCOS SINTÉTICOS .....	39
FIGURA 38: ESQUEMA DE LOS 7 BALONES.....	40
FIGURA 39: IMAGEN ÓPTICA DE LOS 7 BALONES COLGADOS .....	40
FIGURA 40: CORTES EN DISTINTOS PLANOS ACIMUT-ELEVACIÓN (A,B,C) DE LA IMAGEN ACÚSTICA CAPTURADA DE LAS 7 BOLAS .....	41
FIGURA 41: POSICIONES: REAL DE LAS BOLAS (CÍRCULOS ROJOS) Y OFRECIDAS POR EL MODELO (ESTRELLAS AZULES).....	41
FIGURA 42: INFORMACIÓN DEL SOFTWARE DE ADQUISICIÓN SOBRE LOS SENSORES ERRÓNEOS .....	42
FIGURA 43: MEDIDAS DE MEDIA Y DESVIACIÓN TÍPICA EN LA GANANCIA DE CADA UNO DE LOS MICRÓFONOS MEMS PROPORCIONADAS POR EL SISTEMA DE ADQUISICIÓN.....	44
FIGURA 44: COMPARATIVA DE RESULTADOS OBTENIDOS PARA EL EXPERIMENTO PLANTEADO CON (A) EL MODELO SIMPLE Y (B) EL MODELO COMPLEJO. POSICIONES: REAL DE LAS BOLAS (CÍRCULOS ROJOS) Y OFRECIDAS POR EL MODELO (ESTRELLAS AZULES).....	44
FIGURA 45: MEDIDAS ANTROPOMÉTRICAS TOMADAS SOBRE EL INDIVIDUO .....	45

FIGURA 46: ESQUEMA DE LOS PUNTOS SIGNIFICATIVOS DEL INDIVIDUO MEDIDO: (A)VISTA FRONTAL, (B) VISTA LATERAL Y (C) VISTA SUPERIOR .....	46
FIGURA 47: MÁXIMOS CORRESPONDIENTES A CABEZA, PECHO Y CINTURA. (A)VISTA FRONTAL Y (B) VISTA LATERAL .....	47
FIGURA 48: MÁXIMOS CORRESPONDIENTES A BRAZOS Y MANOS (VISTA SUPERIOR) .....	47
FIGURA 49: MÁXIMOS CORRESPONDIENTES A LAS RODILLAS. (A)VISTA FRONTAL Y (B) VISTA LATERAL .....	48
FIGURA 50. MÁXIMOS ASOCIADOS A LÓBULOS SECUNDARIOS. VISTA FRONTAL .....	49
FIGURA 51: HISTOGRAMA DEL RUIDO PRESENTE EN UNA IMAGEN ACÚSTICA .....	51
FIGURA 52: SEÑAL RECIBIDA Y ESTIMACIÓN DEL MODELO PARA UN ÚNICO BLANCO. CASO SIN RUIDO .....	54
FIGURA 53: FUNCIÓN DE COSTES PARA UN ÚNICO BLANCO. CASO SIN RUIDO .....	54
FIGURA 54: SEÑAL RECIBIDA Y ESTIMACIÓN DEL MODELO PARA UN ÚNICO BLANCO CONTAMINADO CON RUIDO EXPONENCIAL DE $\sigma = 6$ .....	55
FIGURA 55: FUNCIÓN DE COSTES PARA UN ÚNICO BLANCO CONTAMINADO CON RUIDO EXPONENCIAL DE $\sigma = 6$ .....	55
FIGURA 56: RESULTADOS DE LA BÚSQUEDA EXHAUSTIVA APLICADA SOBRE EL MODELO DE 3 BLANCOS.....	57
FIGURA 57: MEJOR SOLUCIÓN OBTENIDA PARA 10 ITERACIONES REALIZADAS A TRAVÉS DE LA BÚSQUEDA ALEATORIA FIJANDO EL ERROR CUADRÁTICO MEDIO MÁXIMO EN 0,3 .....	59
FIGURA 58: CONSTRAINED NONLINEAR OPTIMIZATION VI .....	60
FIGURA 59: REPRESENTACIÓN ACIMUT-ELEVACIÓN DE LAS POSICIONES DE LOS 7 BLANCOS SIMULADOS (CRUCES ROJAS) Y ESTIMADAS (PUNTOS DE COLORES) .....	60
FIGURA 60: PARÁMETROS ESTADÍSTICOS DEL ERROR OBTENIDO EN LA ESTIMACIÓN DE CADA BLANCO PARA EL ESCENARIO SIMULADO. (A) MEDIA, (B) DESVIACIÓN ESTÁNDAR.....	61
FIGURA 61: EJEMPLO DE KERNEL 3D (3x3x3).....	62
FIGURA 62: MÁXIMOS CARACTERÍSTICOS OBTENIDOS A PARTIR DE 200 IMÁGENES DE UN INDIVIDUO. (A) VISTA FRONTAL, (B) VISTA LATERAL, (C)VISTA SUPERIOR .....	64
FIGURA 63: INDIVIDUO I1 E INDIVIDUOS I2 E I3 CON SUS RESPECTIVAS VESTIMENTAS (A Y B) .....	66
FIGURA 64: PROBABILIDADES DE DETECCIÓN DE LOS ELEMENTOS DE REFLEXIÓN MEDIANTE DISTINTOS MÉTODOS .....	68
FIGURA 65: ECMs PARA LAS 200 MUESTRAS DEL INDIVIDUO 1 CON LOS PARÁMETROS INICIALES, ESTIMADOS Y CLASIFICADOS TRAS APLICAR LOS CONTENEDORES. ....	70
FIGURA 66: COMPARATIVA DE LAS MÉTRICAS DE EVALUACIÓN GLOBALES PARA CADA MÉTODO APLICADO .....	78
FIGURA 67: PREDICCIONES (%) REALIZADAS A TRAVÉS DEL CLASIFICADOR CONSTRUIDO CON EL MODELO INTELIGENTE DE MÁXIMOS 3D, VALORES DE AMPLITUD ABSOLUTOS Y N=10.....	79

---

## - ÍNDICE DE TABLAS

---

TABLA 1: MODELO SIMPLIFICADO DE 3 BLANCOS 1D PARA UNA PERSONA .....	33
TABLA 2: MODELO SIMPLIFICADO DE 7 BLANCOS 2D PARA UNA PERSONA .....	36
TABLA 3: MODELO SIMPLIFICADO DE 7 BLANCOS 3D PARA UNA PERSONA .....	38
TABLA 4: RESULTADOS OBTENIDOS EN 10 EJECUCIONES DEL ALGORITMO DE BÚSQUEDA ALEATORIA PARA UN ERROR CUADRÁTICO MEDIO MÁXIMO DE 0,3 .....	58
TABLA 5: LÍMITES ESPACIALES DE LOS CONTENEDORES TRIDIMENSIONALES. $\theta^\circ, \varphi^\circ$ y $r$ (cm).....	62
TABLA 6: CARACTERÍSTICAS DE LAS CAPTURAS ACÚSTICAS .....	65
TABLA 7: DISTRIBUCIÓN DEL CONJUNTO DE DATOS DE ENTRENAMIENTO Y DE TEST .....	68
TABLA 8: PROBABILIDADES DE DETECCIÓN DE CADA ELEMENTO DE REFLEXIÓN SEGÚN EL MÉTODO Y NÚMERO DE IMÁGENES PROMEDIADAS (N).....	69
TABLA 9: RESULTADOS DEL PROCESO DE EXTRACCIÓN DE CARACTERÍSTICAS MEDIANTE EL MÉTODO DE OPTIMIZACIÓN .....	69
TABLA 10: RESULTADOS DEL PROCESO DE EXTRACCIÓN DE CARACTERÍSTICAS MEDIANTE EL MODELO INTELIGENTE.....	69
TABLA 11: MATRIZ DE CONFUSIÓN GENERADA UTILIZANDO EL MODELO DE OPTIMIZACIÓN CON PROMEDIOS DE 10 IMÁGENES .....	75
TABLA 12: MÉTRICAS DE EVALUACIÓN OBTENIDAS UTILIZANDO EL MODELO DE OPTIMIZACIÓN CON PROMEDIOS DE 10 IMÁGENES .....	76
TABLA 13: MATRIZ DE CONFUSIÓN GENERADA UTILIZANDO EL MODELO INTELIGENTE CON PROMEDIOS DE 10 IMÁGENES Y VALORES DE AMPLITUD NORMALIZADA.....	76
TABLA 14: MÉTRICAS DE EVALUACIÓN OBTENIDAS UTILIZANDO EL MODELO INTELIGENTE CON PROMEDIOS DE 10 IMÁGENES Y VALORES DE AMPLITUD NORMALIZADA.....	76
TABLA 15: MATRIZ DE CONFUSIÓN GENERADA UTILIZANDO EL MODELO INTELIGENTE CON PROMEDIOS DE 10 IMÁGENES Y VALORES ABSOLUTOS DE AMPLITUD .....	77
TABLA 16: MÉTRICAS DE EVALUACIÓN OBTENIDAS UTILIZANDO EL MODELO INTELIGENTE CON PROMEDIOS DE 10 IMÁGENES Y VALORES ABSOLUTOS DE AMPLITUD .....	77
TABLA 17: MATRIZ DE CONFUSIÓN GENERADA UTILIZANDO EL MODELO INTELIGENTE CON PROMEDIOS DE 20 IMÁGENES Y VALORES ABSOLUTOS DE AMPLITUD .....	77
TABLA 18: MÉTRICAS DE EVALUACIÓN OBTENIDAS UTILIZANDO EL MODELO INTELIGENTE CON PROMEDIOS DE 20 IMÁGENES Y VALORES ABSOLUTOS DE AMPLITUD .....	78

# 1

## Capítulo 1. INTRODUCCIÓN

---

¿Cuántas veces al día utiliza una persona promedio su huella dactilar o el reconocimiento facial para desbloquear su teléfono móvil? Sin duda, la identificación biométrica cada vez está más presente en nuestras vidas, ya sea a través del uso de la huella dactilar, rasgos faciales o voz entre otros [1].

Sin embargo, hemos asimilado tanto estos sistemas biométricos que ninguno de ellos es ya sorprendente para nosotros. Ahora bien, imagine escuchar un pitido y poder ser reconocido a través de los ecos que su cuerpo refleja sobre él. Esto puede parecer difícil de creer, pero, de hecho, algunas especies de murciélagos son capaces de identificar distintos tipos de flores mediante estas técnicas pulso-eco [2].

A raíz de la pandemia global causada por la enfermedad COVID-19, una de las medidas más comunes empleada por los gobiernos del mundo fue el uso de una mascarilla protectora. Llevar mascarilla oculta información clave para el reconocimiento facial, por lo que decrece la precisión del sistema [3]. Asimismo, otros sistemas como el reconocimiento por huella dactilar implicaban “tocar” de manera directa el sistema y esto supone un problema teniendo en cuenta que el virus puede ser transmitido por contacto [4]. Además, algunos estudios corroboran que el funcionamiento de los sistemas biométricos multimodales es muy superior a los monomodales, por lo que buscar nuevos métodos de reconocimiento supone una idea prometedora [1], [5].

Cada uno de estos sistemas de identificación funcionan esencialmente como un sistema de reconocimiento de patrones. Es decir, consiguen extraer de un conjunto de datos una serie de parámetros o características que identifiquen de manera inequívoca al individuo en cuestión [1]. De esta manera, al igual que las líneas de las yemas de los dedos identifican a una persona, los ecos acústicos reflejados por cada persona deberían ser suficientemente representativos de su identidad, como se ha confirmado en investigaciones anteriores [6].

Para este tipo de problema, los algoritmos de aprendizaje automático son una solución recurrente. Sin embargo, los datos originales contienen mucha información redundante y poco fiable, por lo que debe ser preprocesada. Así, se extraen y seleccionan las características principales de manera inteligente hasta llegar a lo que se conoce como *training set*. De hecho, la preparación de los datos es la parte más importante a la hora de enfrentar un problema de *machine learning* [7].

Esta extracción de características requiere un estudio previo de las imágenes. La idea es plantear un modelo suficientemente representativo de las personas, que recoja los principales patrones de las imágenes para que las características sean significativas.

Además, antes de abordar el problema del preprocesado, se deben adquirir las imágenes a través de un array acústico. La disposición, frecuencia de pulso y otros parámetros de conformado determinarán la calidad para fines biométricos de las imágenes obtenidas, por lo que será conveniente elegir de manera correcta los valores de dichas variables.

Finalmente, hay que recalcar que el sistema de identificación biométrica que se plantea, al estar basado en reflexiones acústicas, tiene una gran ventaja en entornos de poca o nula visibilidad. Mientras que una cámara de video resultaría inútil, el sistema acústico respondería de igual manera puesto que el sonido no precisa de luz para su propagación. Esta característica dota al sistema de posibles aplicaciones interesantes como la vigilancia o identificación en entornos de trabajo de baja visibilidad. Además, se trata de un sistema de bajo coste, fácilmente escalable y no intrusivo.

## **1.1 OBJETIVOS**

### ***1.1.1 OBJETIVO GENERAL***

A partir de imágenes acústicas adquiridas de manera inteligente mediante un array de sensores MEMS de alta resolución y a través de LabVIEW, se pretende extraer una serie de características a través de un modelo que asocie a cada imagen un conjunto de parámetros. Utilizando dichas características que dependerán principalmente de la figura antropométrica del individuo, se clasificarán distintas personas utilizando técnicas de aprendizaje automático supervisado.

### ***1.1.2 OBJETIVOS ESPECÍFICOS***

A mayores de la consecución del objetivo general, se han considerado una serie de objetivos particulares:

- Decidir los parámetros y posición óptima del sistema de adquisición para optimizar la calidad de las imágenes acústicas para biometría.
- Estudiar la dependencia de las imágenes acústicas con la figura antropométrica humana y otros factores.
- Modelar y cuantificar el efecto del ruido presente en las imágenes acústicas obtenidas.
- Proporcionar un modelo sintético y verosímil de una figura humana.
- Plantear una estrategia de estimación paramétrica válida
- Entrenamiento de un algoritmo de *machine learning* para caracterizar las prestaciones del clasificador según el modelo planteado.
- Proporcionar un análisis estadístico final del identificador biométrico
- Recoger las dificultades y consideraciones más significativas de la tecnología de identificación.

## **1.2 ESTRUCTURA Y FASES DEL PROYECTO**

En base a la consecución de los objetivos planteados y teniendo en cuenta las fases del proyecto, se presenta y justifica a continuación la estructura que sigue la memoria.

Primeramente, en el marco teórico, se entrega una perspectiva al lector acerca del proyecto planteado. Para ello se lleva a cabo una explicación y análisis de distintos conceptos considerados como necesarios para interpretar los resultados, afirmaciones y suposiciones que se plantean a lo largo del estudio.

Acto seguido, los capítulos posteriores se organizan abordando el proyecto de manera secuencial. Las fases principales del proyecto se definen a continuación:

### ***1.2.1 ADQUISICIÓN DE IMÁGENES ACÚSTICAS***

Como se ha justificado previamente, el proyecto comienza desde la adquisición de las imágenes acústicas. En el capítulo 3 de la memoria se presentará el sistema de adquisición, tanto hardware como el software SAM (Sistema de Adquisición MEMS).

Después, se decidirá la disposición de los sensores en el array, así como su posición física óptima, parámetros en la conformación y otras variables significativas como el tiempo de pulso transmitido, las frecuencias de transmisión o el margen de captura y conformación.

También, se presenta el escenario de adquisición y se realizan distintas consideraciones sobre la figura antropométrica humana y sobre la construcción de las imágenes acústicas a partir de la información recibida por los sensores que permitan entender su origen adecuadamente. Por último, también se comentan en este capítulo algunas técnicas simples de preprocesado que faciliten la diferenciación entre los blancos reales y el *clutter*.

### ***1.2.1 DISEÑO DE UN MODELO PARAMÉTRICO PARA BIOMETRÍA***

Con el fin de reducir el volumen de información de las imágenes acústicas, en el capítulo 4 se desarrollará un modelo paramétrico. La idea principal es basar el modelo en las respuestas espaciales del array 2D para analizar cada imagen acústica como una combinación lineal de respuestas espaciales para los blancos presentes en una imagen.

Particularizando este modelo para las imágenes de biometría, aunque se discute más adelante, se asume de partida que una persona tiene un número finito de elementos de reflexión: (cabeza, pecho, brazos y piernas).

De esta manera, según el modelo planteado, una imagen de una persona se podrá caracterizar como una combinación lineal y finita de blancos puntuales y a su vez cada uno de ellos como una cuaterna de parámetros (ganancia, acimut, elevación y rango).

### ***1.2.2 EXTRACCIÓN DE PARÁMETROS***

De manera consistente con el modelo escogido previamente, es preciso extraer los parámetros correspondientes a las imágenes acústicas. El capítulo 5 aborda la estimación de estas características necesarias para el entrenamiento del modelo.

Se plantean las ventajas e inconvenientes de basar la estimación en métodos heurísticos en contraposición con el método de estimación clásica LSE (*Least Squares Estimator*).

Es decir, el objetivo aquí será extraer consecuentemente con el modelo escogido, de las imágenes acústicas obtenidas, un conjunto de parámetros que sean óptimos y suficientemente representativos para poder posteriormente realizar su clasificación.

### **1.2.3 FASE DE ENTRENAMIENTO Y CLASIFICACIÓN**

Una vez confeccionado por completo el sistema de extracción de características, el último paso será definir e implementar el algoritmo de *machine learning*. Esto se lleva a cabo en el capítulo 5 donde se tratan las fases finales de entrenamiento y clasificación de las imágenes.

Para ello, previamente se ha creado una base de datos que cuente con múltiples imágenes acústicas de diferentes personas. Seguidamente, se entrena al sistema con los diferentes parámetros obtenidos a partir de varias imágenes de diferentes personas para confeccionar así un modelo.

Finalmente, para comprobar el funcionamiento del sistema, se introducen imágenes nuevas (es decir, que no se hayan utilizado ya en la fase de entrenamiento) de personas con las que ya se haya entrenado al sistema para identificar. El objetivo final es proporcionar un análisis de la probabilidad de acierto y error del sistema a la hora de realizar la clasificación teniendo presente la aplicación como identificador biométrico del sistema.

### **1.2.4 CONCLUSIONES Y VALIDACIÓN FINAL**

Tras haber ejecutado debidamente todas las fases del proyecto, en el último capítulo de la memoria, se explica y analizan las conclusiones y conflictos que a lo largo del proyecto se han ido planteado, así como ciertas consideraciones sobre las posibles líneas de mejora del sistema.

## **1.3 ESTADO DEL ARTE**

En la actualidad aún es muy escasa la investigación sobre sistemas de reconocimiento biométrico basados en técnicas pulso-eco por medio de un array acústico. Sin ir más lejos, el uso de las técnicas pulso-eco es mucho más habitual en el ámbito de la detección. De hecho, las tecnologías RADAR y SONAR siempre han estado muy presentes tanto en la práctica como en la investigación en la localización de obstáculos en navegación o conducción.

Fuera del ámbito de identificación biométrica, los estudios de Moebus y Zoubir, abordan el problema de la parametrización de imágenes obtenidas a través de un array acústico modelando los ecos a través de GMM (*Gaussian-Mixture-Model*) [8]. Estos mismos autores emplean una banda de frecuencia ultrasónica para generar a través de un array planar imágenes acústicas 3D usando técnicas de conformado (*beamforming*) [9]. De hecho, Moebus aplicó estos avances en su tesis donde consiguió clasificar distintos

objetos a través de sus imágenes acústicas para poder discriminar humanos de otros objetos con una precisión de casi 97% [10].

Por otra parte, el uso de algoritmos de aprendizaje automático como SVM se encuentra bastante extendido en el campo de la identificación biométrica independientemente del origen de los datos [11]–[13].

Por último, en estudios previos a esta investigación, se propuso un sistema de biometría acústica. El sistema utilizaba un array unidimensional, preprocesaba las imágenes utilizando técnicas GMM y por último las clasificaba a través de SVM (*Support Vector Machine*) [14].

## Capítulo 2. MARCO TEÓRICO

### 2.1 IMÁGENES ACÚSTICAS MULTIFRECUENCIA

Utilizando un array de micrófonos 2D y enviando una señal sinusoidal pulsada a través de un transmisor, se obtiene una imagen acústica en 3 dimensiones. Si además se tiene en cuenta que el sistema puede funcionar en múltiples frecuencias, se puede considerar estas imágenes como 4D.

La representación gráfica de las características de radiación de una antena es el diagrama de radiación (*beam pattern*). A través de esta representación, se puede saber por ejemplo la dirección de apuntamiento, la directividad de la antena o el ancho del lóbulo principal. El diagrama de radiación de la antena se puede representar en tres o en dos dimensiones dependiendo de la dimensionalidad del espacio de estudio. Así, si se quiere únicamente ver la respuesta de la antena ante una coordenada, será suficiente con una representación 2D como las siguientes (*Figura 1*).

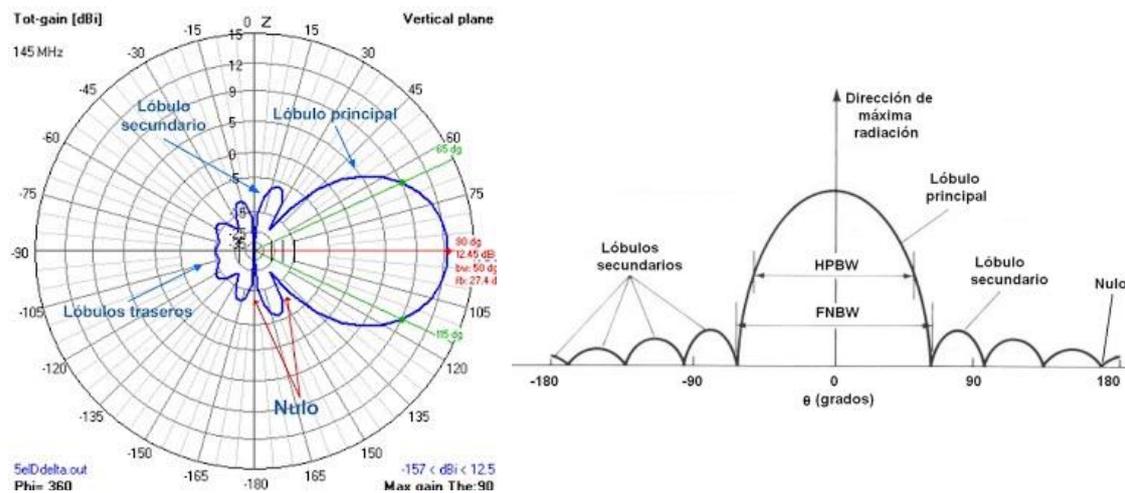


Figura 1: Representaciones 2D del diagrama de radiación de una antena

El principio pulso-eco (*Figura 2*) permite saber a qué distancia se encuentra un objeto para una dirección espacial determinada si el diagrama de radiación de la antena es lo suficientemente estrecho.

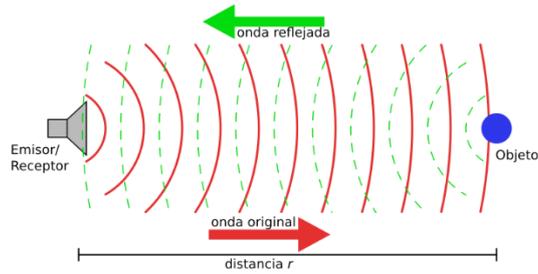


Figura 2: Técnica pulso-eco aplicada con un solo transceptor

Sin embargo, una imagen acústica se obtiene cuando se escanea un área del espacio. Es decir, se barren distintos ángulos que son abarcados en la región objetivo. Este escaneo se puede realizar de forma mecánica o electrónica. Las antenas de barrido mecánico suelen incluir una o más piezas articuladas como lentes o espejos que permiten al sistema apuntar físicamente a las distintas zonas del espacio para así capturar la región de interés [15].

Al disponer de un array de sensores, se puede realizar este escaneo de forma electrónica gracias a un proceso de conformación de haces (*beamforming*). De este modo, es posible construir una imagen acústica a través de la información recibida en cada uno de los sensores del array.

Un transmisor genera la señal pulsada y se forma una imagen a partir de las señales provenientes de los ecos que se reciben en cada uno de los distintos micrófonos del array [16]. Para determinar los ángulos y distancia a la que se encuentra el blanco, se siguen los principios básicos de los sistemas RADAR (*Radio Detection and Ranging*). Se calcula la distancia a partir de la velocidad del sonido en el medio y el tiempo empleado por la onda acústica en viajar hasta el blanco y volver. La dirección o posición angular se puede determinar gracias a la dirección de llegada de la onda reflejada [17].

El conformador (*beamformer*), como se puede ver en la figura inferior (*Figura 3*), combina linealmente las señales recibidas en cada micrófono  $x_n$ . Para ello, las multiplica por sus pesos complejos asociados  $w_n$  y se suman entre sí obteniendo una señal de salida correspondiente a la imagen acústica [6]. Así, modificando el valor de los desfases de estos pesos, se consigue que el diagrama de radiación “apunte” a posiciones espaciales diferentes, ya que según los valores de  $w_n$ , las señales capturadas por los micrófonos se sumarán o no en fase.

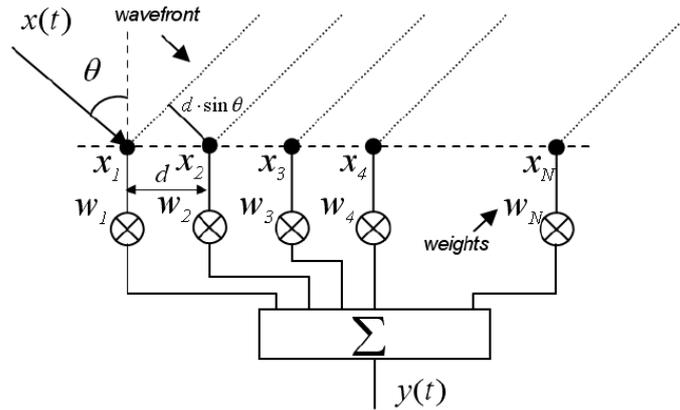


Figura 3: Estructura de un conformador [6]

Existen distintas técnicas de conformado y de hecho se puede reducir el nivel de los lóbulos secundarios en el diagrama de radiación modificando el módulo de los pesos complejos  $w_n$ . Esta práctica se conoce como enventanado espacial y puede ser muy útil dependiendo la aplicación. Pero no todo son ventajas ya que disminuir el efecto de los lóbulos secundarios implica un aumento del ancho del lóbulo principal y por ende una pérdida de resolución espacial. Enventanar también permite tener en cuenta los valores de ganancia individual de cada uno de los micrófonos para normalizar sus diferencias.

Estas imágenes se obtienen en sistema de coordenadas esféricas por lo que las 3 primeras dimensiones representan acimut, elevación y rango, mientras que la cuarta representaría, para las imágenes multifrecuencia, cada una de las frecuencias utilizadas.

El centro del array corresponde al centro de coordenadas  $O$ , el rango  $r$  es la distancia desde ese centro hasta el punto y por último las magnitudes acimut  $\theta$  y elevación  $\varphi$ , recogen los ángulos horizontal y vertical respectivamente (Figura 4).

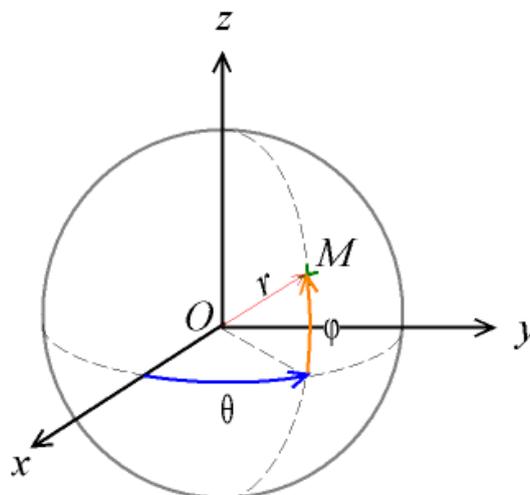
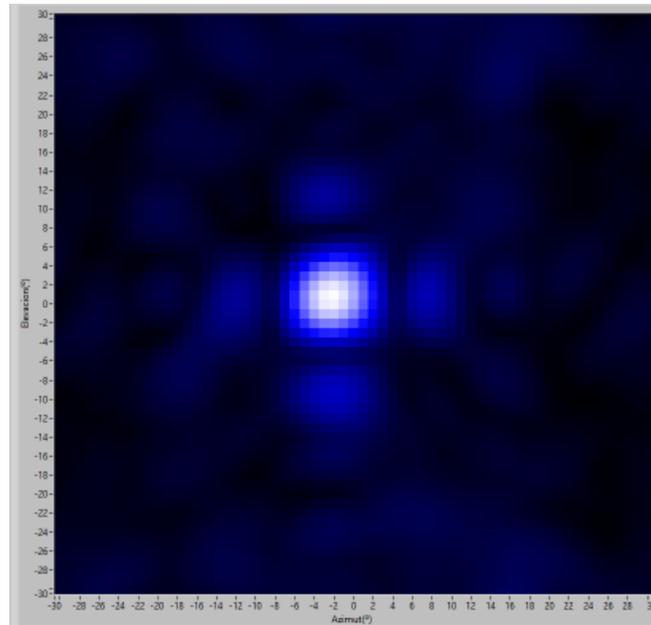


Figura 4: Sistema de coordenadas esféricas

Para visualizar las imágenes 3D, típicamente se utilizarán proyecciones sobre las distintas coordenadas o en su defecto cortes en la imagen. Además, en preprocesado, las imágenes se normalizan respecto a su valor máximo de amplitud por lo que únicamente toman valores entre 0 y 1. A continuación, se muestra un ejemplo de una imagen acústica de un único blanco y representada por su proyección sobre el plano formado por las coordenadas acimut y elevación (*Figura 5*). Los puntos con mayor energía se representan como puntos más claros y los que menos energía contienen se representan en un tono más oscuro.



*Figura 5: Proyección sobre el plano acimut-elevación de un blanco puntual*

## 2.2 BIOMETRÍA ACÚSTICA Y ANTROPOMETRÍA

El término “biometría” se refiere al hecho de identificar una persona en base a sus características fisiológicas o comportamiento. Un sistema biométrico adquiere datos biométricos de un individuo, extrae las características consideradas y posteriormente compara dichas características con un modelo existente creado previamente. Así, se decide o bien quién es la persona en cuestión o bien si es quien dice ser dependiendo si el sistema está orientado a identificación o verificación de individuos [18].

Existen varias cualidades que se buscan a la hora de escoger unas características biométricas de calidad. Principalmente: universalidad, singularidad, permanencia, desempeño, aceptabilidad, y fiabilidad [19]. Las cualidades más relevantes para este trabajo serán las 4 primeras:

- Universalidad se refiere al hecho de encontrar unos factores biométricos que estén presentes en todos los individuos.
- Singularidad es una medida de cómo de único es ese factor biométrico sobre la población total de los individuos.

- Permanencia se refiere a cómo de constante se mantiene la característica correspondiente con el paso del tiempo.
- Por último, desempeño se refiere a cómo de bien funciona el sistema. Factores como la exactitud, rapidez y robustez son tenidas en cuenta acorde con el requerimiento de recursos.

El término “biometría acústica” se refiere a las técnicas de biometría que se basan en ondas acústicas para su funcionamiento. En el caso en el que la frecuencia de las ondas se encuentre por encima del espectro auditivo (aproximadamente 20 kHz), se suele emplear el término “biometría ultrasónica”.

Por otra parte, asumiendo que el factor principal que origina las diferencias entre las imágenes acústicas de unas personas y otras es la figura antropométrica del individuo, cabe preguntarse cómo de diferentes son unas figuras humanas de otras. Es por ello por lo que la antropometría tiene cabida en este estudio.

Antonio Carmona, recoge en su libro [20] los principales aspectos antropométricos de la población laboral española. En este libro, se realiza un análisis estadístico de diversas medidas tomadas a una serie de individuos.

De los datos presentados en el mencionado escrito, se puede inferir que los percentiles en los que una persona se encuentra para las diferentes medidas antropométricas son notablemente distintos entre sí. Es evidente que un individuo que presente una alta estatura contará seguramente con una gran envergadura, pero difícilmente se encontrará en el mismo percentil con respecto al resto de la población en ambas medidas.

Esto es una conclusión interesante, ya que demuestra la utilidad de considerar diversas medidas de un individuo para su identificación. Esto no sería productivo si las distintas medidas fuesen redundantes entre sí.

Finalmente, la respuesta acústica asociada a las características antropométricas de cada individuo es lo que se denominará como “firma acústica” y será identificativo de cada persona.

### **2.3 ESTIMACIÓN, OPTIMIZACIÓN Y HEURÍSTICA**

La teoría de la estimación es el elemento clave en gran parte de los sistemas electrónicos de procesamiento de señal diseñados para extraer información. Sin ir más lejos, los sistemas RADAR y SONAR (*Sound Navigation And Ranging*). Ambos se basan en el cálculo del retardo de ida y vuelta de los pulsos transmitidos. No obstante, en aplicaciones reales, los pulsos llegan atenuados o contaminados con distintas fuentes de ruido aleatorio. Es decir, en los sistemas reales hay diversos problemas que requieren de técnicas de estimación para obtener el valor de parámetros clave [21].

Aún más cercano al problema de identificación en estudio es el ejemplo que utiliza Kay sobre el reconocimiento de vocales. Teniendo almacenada la señal acústica de cada vocal pronunciada por una persona, se puede identificar la vocal a la que corresponde un sonido

nuevo. Básicamente, se compara con la base de datos y se escoge la vocal que más “se parece” al sonido recibido minimizando alguna medida de diferencia [21].

De esta manera, resolver el problema inicial puede pasar por la resolución de un problema de optimización. Así pues, los métodos de optimización clásicos como LP (*Linear Programming*) o NLP (*Non Linear Programming*) permiten encontrar el óptimo de la función objetivo satisfaciendo el conjunto de restricciones que requiera el problema [22].

Existen algunos conceptos básicos que se deben conocer para todo problema de optimización [22]:

- La región factible representa todos los puntos susceptibles de ser una solución al problema y es un importante elemento a tener en cuenta.
- La función objetivo representa una medida del rendimiento del sistema y dependiendo del objetivo, se puede buscar encontrar su máximo o mínimo absoluto.
- Las variables de optimización representan las decisiones que hacen variar el valor de la función objetivo y sobre las que se plantea la optimización.
- Si se trata de un problema con restricciones, estas representan el conjunto de ecuaciones o desigualdades que deben cumplirse para considerar una solución como válida.

Por otra parte, el uso de algoritmos heurísticos conlleva soluciones sin garantía de optimalidad pero que suelen ejecutarse con mayor rapidez [22]. Según Zanakis y Evans, para los profesionales, los métodos heurísticos son simples procedimientos, normalmente guiados por el sentido común, que proporcionan soluciones buenas, pero no necesariamente óptimas, para problemas difíciles y de manera fácil y rápida [23].

Por último, cabe mencionar que un problema de estimación puede resolverse mediante un enfoque clásico, considerando el parámetro a estimar como una constante determinista, o bayesiano, modelando el parámetro como una variable aleatoria si se conoce información a priori [21].

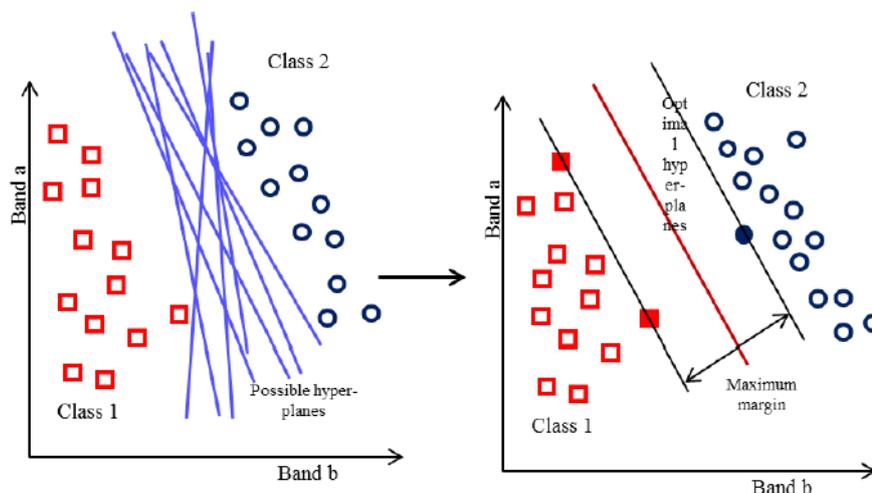
## 2.4 MACHINE LEARNING. SVM

Existen numerosas situaciones donde no se puede especificar con precisión un método por el cual se puede calcular la salida correcta a partir de las entradas. Estas tareas no pueden resolverse con un enfoque tradicional por lo que requieren de otras estrategias alternativas [24].

Una de estas ideas alternativas es hacer que el algoritmo intente aprender la funcionalidad de entrada/salida a partir de ejemplos, del mismo modo que los niños aprenden por ejemplo que coches son deportivos simplemente diciéndoles cuales de un gran número de coches son deportivos en lugar de recibir una explicación precisa sobre las cualidades que hacen a un coche ser deportivo [24].

Este enfoque basado en utilizar ejemplos para obtener algoritmos se conoce como metodología de aprendizaje o *machine learning*, y en el caso concreto en el que los ejemplos sean pares de entrada/salida se denomina aprendizaje supervisado [24].

Uno de los algoritmos más utilizados es la máquina de vectores de soporte o SVM (*Support Vector Machines*). Este algoritmo proporciona una excelente separación de clases cuando se utiliza como clasificador supervisado [25]. SVM se basa en determinar los hiperplanos que separan de forma óptima los puntos pertenecientes a distintas clases (*Figura 6*). El objetivo es maximizar los márgenes entre clases para alcanzar el clasificador más eficiente [25]. El vector formado por los puntos más cercanos al hiperplano recibe el nombre de vector de soporte (círculo y cuadrados coloreados en la (*Figura 6*)).



*Figura 6: Problema de clasificación binaria en 2 dimensiones a través de SVM [25]*

Cada una de las categorías o valores de salida posibles del sistema clasificador se denomina como “clase”. Si el sistema cuenta con más de 2 clases, el problema pasa de ser un problema de clasificación binaria a uno de clasificación multiclase. Estos casos se resuelven asociando un hiperplano a cada clase. De esta manera los algoritmos aprenden simultáneamente todos los hiperplanos a partir de los datos [24].

En muchas ocasiones, el conjunto de datos puede estar desbalanceado, es decir, hay alguna clase que cuenta con muchas menos muestras en relación con el resto. Esto puede hacer que el sistema se incline por alguna clase generando un sesgo indeseado en el algoritmo [7]. En este caso sería necesario equilibrar los datos para obtener una muestra más balanceada.

Por otra parte, las cualidades que se extraen del conjunto de los datos “en crudo” para entrenar al modelo se conocen como características. La obtención de estas características es uno de los pasos más importantes, aunque no hay que subestimar la importancia de separar correctamente los datos disponibles en el conjunto de entrenamiento (*training dataset*), de validación (*validation dataset*) y de prueba (*test dataset*) [25].

Una vez separados, el procedimiento seguido para construir el clasificador se puede ver en el esquema inferior (*Figura 7*). A partir de los datos de entrenamiento se genera un clasificador que se evalúa con las nuevas muestras que fueron destinadas a validación. Así, se puede averiguar y perfeccionar el rendimiento del sistema comparando las etiquetas reales con las predichas por el modelo acorde con el criterio decidido [14].

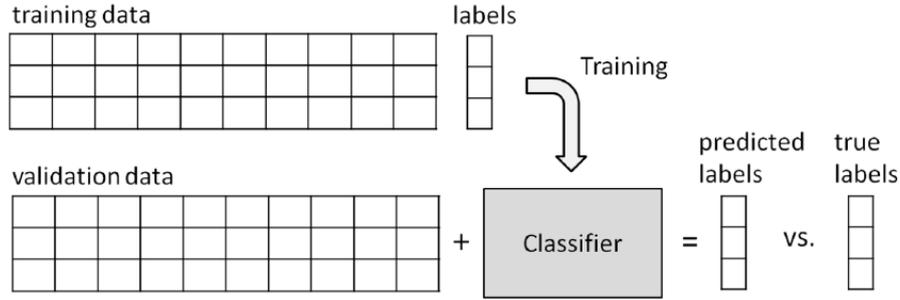


Figura 7 Entrenamiento y validación del identificador. [14]

Normalmente, se quiere utilizar la mayor cantidad posible de los datos disponibles para el entrenamiento. Sin embargo, si el conjunto de datos destinados a validación es pequeño, la estimación del rendimiento del modelo no será demasiado precisa [26].

Una solución a este problema pasa por utilizar la técnica de validación cruzada conocida como SFCV (*S-Fold Cross Validation*). A continuación, se ilustra un ejemplo de SFCV para  $S = 4$  (Figura 8). Esta técnica consiste en dividir los datos disponibles en  $S$  grupos y utilizar  $S - 1$  grupos (coloreados en blanco) para entrenar unos modelos que se evalúan en el grupo restante (coloreado en rojo). El procedimiento se repite para las  $S$  combinaciones posibles y las puntuaciones de rendimiento de las  $S$  ejecuciones se promedian [26].

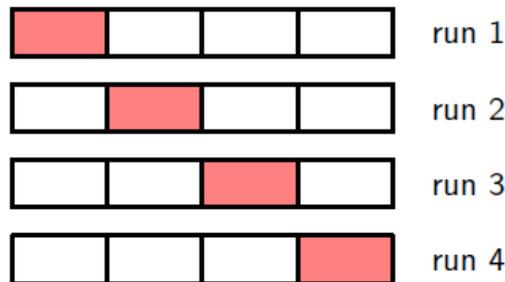


Figura 8: Técnica de validación cruzada SFCV con  $S = 4$  [26]

La principal desventaja de esta técnica es que el número de ejecuciones de entrenamiento se incrementa en un factor  $S$  lo que resulta en un aumento de la carga computacional [26].

En la fase de validación, existen diversas métricas que permiten valorar el rendimiento de los algoritmos. La manera de evaluar el rendimiento del algoritmo por excelencia es la matriz de confusión (Figura 9). Cada celda de la tabla representa el número de veces que el modelo ha predicho correcta o incorrectamente cada clase y permite ver cómo son las equivocaciones [25]. De hecho, se llama “matriz de confusión” porque hace que sea fácil detectar dónde el sistema está confundiendo dos clases.

		Predicted Values			
		$C_1$	$C_2$	...	$C_n$
Actual Values	$C_1$	$N_{11}$	$N_{12}$	...	$N_{1n}$
	$C_2$	$N_{21}$	$N_{22}$	...	$N_{2n}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$C_n$	$N_{n1}$	$N_{n2}$	...	$N_{nn}$

Figura 9: Matriz de confusión para clasificadores multiclase [25]

A partir de los valores contenidos en esta tabla se pueden calcular varias métricas que dan idea del rendimiento del sistema como la precisión, exhaustividad, o el valor-F [25].

Por último, siempre hay que intentar evitar los problemas de sobreajuste (*overfitting*) o subajuste (*underfitting*) (Figura 10). El primero ocurre cuando el modelo de aprendizaje es capaz de ajustarse al 100% de los datos de entrenamiento y hace que el modelo no pueda ser generalizado para otros datos. El segundo ocurre cuando el modelo no predice correctamente los datos de entrenamiento después de haber sido entrenado con ellos. Ambas situaciones llevan a un clasificador de bajo rendimiento por lo que se deben intentar evitar. Para evitar el *overfitting*, se pueden reducir las características o aumentar los datos de entrenamiento, mientras que para combatir el *underfitting*, una solución es aumentar las características [25].

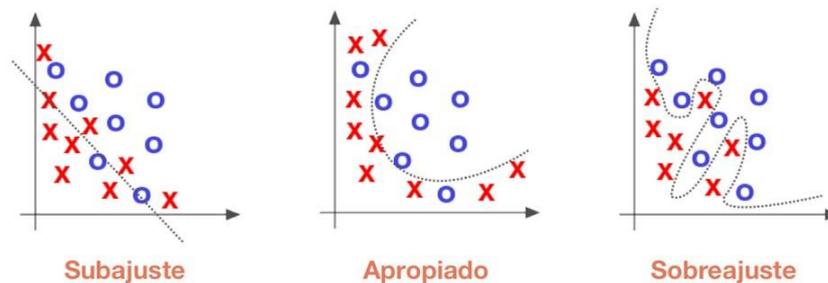


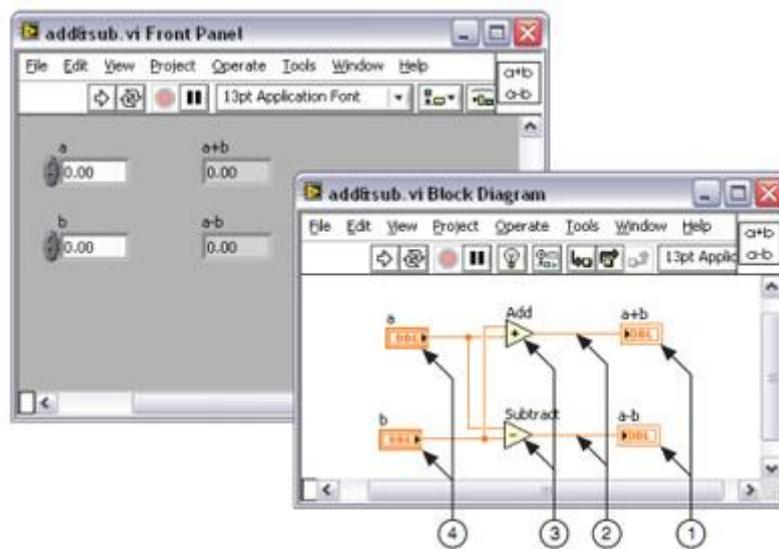
Figura 10: Problemas principales en el aprendizaje supervisado

## 2.5 ENTORNO DE DESARROLLO LABVIEW

Jeff Kodosky, cofundador de NI (*National Instruments*) en 1976, es considerado como el padre de LabVIEW por científicos de todo el mundo [27]. Cuando en 1986 se lanzó la primera versión de LabVIEW, fue anunciada como una manera no programática de crear software para pruebas y medidas. Sin embargo, detrás de esto, existía un lenguaje de programación gráfico riguroso y bien definido conocido como “G” [28].

LabVIEW es un entorno de programación gráfico, es decir, se representa como un diagrama bidimensional formado por nodos y cables interconectados entre sí. El flujo de datos indica la semántica del lenguaje. Es decir, los nodos realizan cálculos basados en sus valores de entrada y sólo cuando se completan propagan sus valores de salida hacia los nodos posteriores. De esta manera, cada cable transporta un único dato que puede ser de distintos tipos (entero, flotante, cadena de caracteres, booleano...) [28].

Cada módulo de software en LabVIEW/G recibe el nombre de instrumento virtual o “VI”. Al abrir un VI, aparecen dos ventanas (*Figura 11*): el panel frontal y el diagrama de bloques. El panel frontal corresponde a la interfaz de usuario mientras que el diagrama de bloques contiene todos los nodos y cables que transfieren datos intercomunicándose con otros objetos del diagrama de bloques [29].



*Figura 11: Panel frontal y diagrama de bloques [29]*

Existen numerosos tipos de nodos. Los terminales de indicador y control (1 y 4 en la imagen superior) permiten introducir y visualizar datos a través del panel frontal. También puede haber funciones, constantes, estructuras o subVIs. Estos últimos corresponde a una llamada de subrutina y permiten ejecutar un VI dentro de otro [29].

A través de las distintas paletas de Controles y Funciones, se puede acceder a los distintos módulos previamente instalados en el PC [29]. Por ejemplo, en la categoría “Optimization VIs” se puede acceder a varias funciones de optimización. En la imagen inferior (*Figura 12*), se puede ver la paleta de controles con la categoría Moderna desplegada.

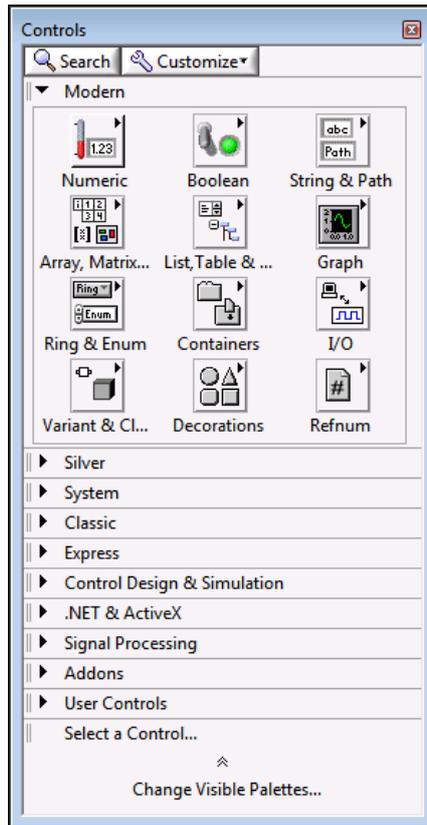


Figura 12: Paleta de controles

# 3

## Capítulo 3. ADQUISICIÓN DE IMÁGENES

---

Como se anticipó en la introducción, adquirir unas buenas imágenes acústicas es fundamental para lograr una buena clasificación final. Para ello, es importante hablar del sistema de adquisición hardware y de las estrategias y parámetros más significativos que intervienen en la conformación y el preprocesado. Por último, es importante también realizar ciertas consideraciones sobre el escenario de adquisición o la postura de la persona.

### 3.1 SISTEMA DE ADQUISICIÓN HARDWARE

El sistema de generación de imágenes acústicas es en esencia el que se describe en el artículo “*Design and Evaluation of a Scalable and Reconfigurable Multi-Platform System for Acoustic Imaging*” [30]. Principalmente, se puede resumir en 3 bloques principales que funcionan de manera coordinada entre sí:

#### 3.1.1 TWEETER

Un altavoz especializado en altas frecuencias (*tweeter*) es el encargado de generar la señal acústica a la frecuencia deseada. Este altavoz transmite los pulsos acústicos según el ancho de pulso y la frecuencia configurada en el software.

El hecho de utilizar un único altavoz implica que no se practica ningún tipo de conformado en transmisión. Utilizar  $N$  altavoces permitiría utilizar un pulso con una intensidad acústica  $N$  veces mayor. De esta manera, el sistema sería capaz de concentrar mucha energía en un punto, pero esto obligaría a utilizar mucho más tiempo para escanear una zona del espacio ya que habría que ir apuntando uno a uno con el array de altavoces a cada punto del espacio. Por el contrario, el conformado aplicado en recepción permite muestrear una única vez a todos los sensores en paralelo por lo que no aumenta el tiempo de adquisición.

Aunque finalmente se ha utilizado un *tweeter* pequeño ubicado en el centro del array (*figura 13*), al principio se probó a utilizar un altavoz externo que se encontraba situado debajo del array (*figura 14*).

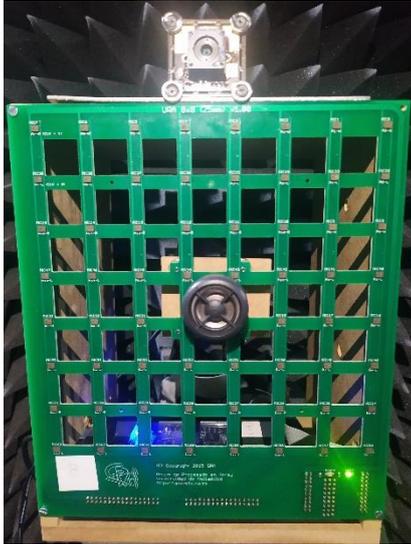


Figura 13: Disposición definitiva del sistema de adquisición



Figura 14: Primera disposición del sistema de adquisición

El problema con la primera disposición (figura 14) se encuentra en la diferencia de los caminos que recorre la onda acústica. Aunque se podría tener en cuenta, el software de adquisición estima la distancia en rango a la que se encuentra el blanco como  $d/2$ , siendo  $d$  el total de distancia recorrida por la onda. Es decir,  $d = d_1 + d_2$  pero  $d_1 \neq d_2$  (figura 15).

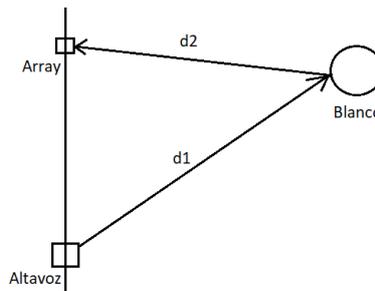


Figura 15: Diferencia de caminos entre las ondas incidente y reflejada

De este modo, si la diferencia entre los caminos que recorre la onda incidente y la reflejada es grande, la calidad de la estimación empeora puesto que el sistema considera que son iguales. Esta es la razón por la que se opta por una disposición donde la distancia entre los micrófonos y el altavoz es mínima.

Por último, hay que tener en cuenta que en la disposición definitiva el tweeter cubre los 4 micrófonos centrales lo que obliga a desactivarlos, puesto que no pueden recibir las ondas reflejadas.

### 3.1.2 MEMS ARRAY

El sistema encargado de capturar los ecos acústicos reflejados consiste en un array UPA (*Uniform Planar Array*) formado por 64 (8x8) micrófonos MEMS que se encuentran uniformemente espaciados 2.5 cm entre sí en una PCB rectangular (Figura 16).

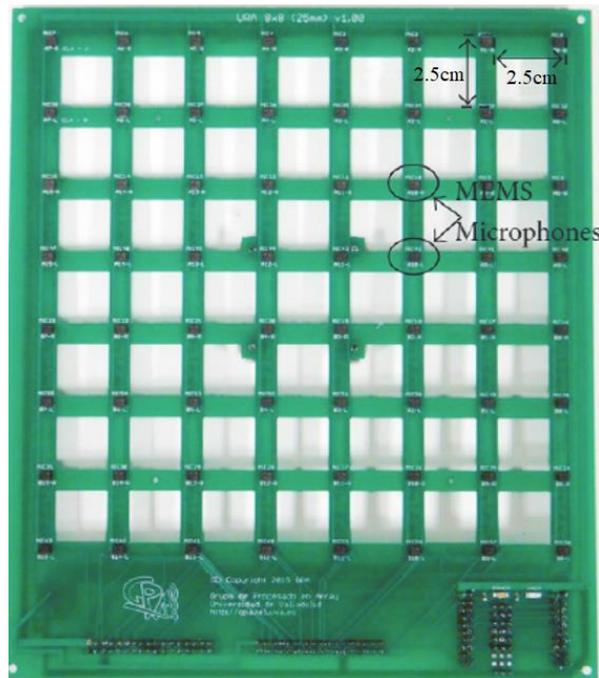


Figura 16: MEMS array

Los micrófonos digitales MEMS (*Micro-Electro-Mechanical System*) están compuestos por dos componentes principales, el sensor acústico y el circuito controlador que incluirá un convertidor ADC. A parte de ser una tecnología barata, la tecnología MEMS reduce significativamente el área ocupada por el sistema lo que hace posible construir arrays con un mayor número de sensores [30], [31].

El sensor normalmente es un elemento capacitivo formado por dos placas de silicio, una fija, llamada *backplate* y otra flexible, llamada *diaphragm*. Los cambios de presión en el aire generados por las ondas acústicas hacen moverse a la membrana flexible. Cuando esto ocurre, la capacitancia entre la placa móvil y la fija cambia y el circuito controlador mide los valores obtenidos [31].

Los micrófonos MEMS se encapsulan para proteger el circuito de agentes externos y permitir a las ondas de sonido llegar al sensor acústico sólo a través de la apertura en la parte superior o inferior del sensor [31].

Para este proyecto se han utilizado los micrófonos MEMS MP34DT01. Observando la hoja de especificaciones técnicas del producto [32], se puede ver como el fabricante (STMicroelectronics) promete una sensibilidad casi constante de 100 a 5000 hercios (*figura 17*). Aunque no forma parte del rango de frecuencias representadas en la gráfica, se puede intuir por el trazado un aumento en la sensibilidad del sistema electrónico para frecuencias mayores como los 14Khz que se utilizan en este estudio.

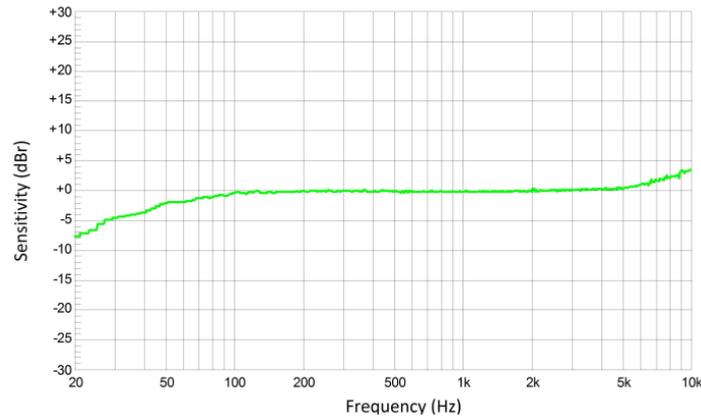


Figura 17: Respuesta frecuencial de los micrófonos en el rango 20Hz - 10KHz

### 3.1.3 SISTEMA DE PROCESADO

La clave del sistema de adquisición de imágenes es el sistema de NI myRIO [33]. Esta plataforma pertenece a la familia de dispositivos reconfigurables Entrada-Salida o RIO (*Reconfigurable Input-Output*) de NI. Estos dispositivos están orientados a sensores cuyos procedimientos de adquisición no estén estandarizados [30].

El procesador de myRIO es capaz de ejecutar a la vez todos los algoritmos software para generar las imágenes acústicas. Además, la plataforma puede almacenar las imágenes en la memoria interna o en un disco externo conectado a través del puerto USB. Sin embargo, aunque myRIO podría funcionar como un sistema aislado, al no contar con un sistema de visualización, se suele controlar con un PC a través de su interfaz WiFi [30].

La myRIO cuenta con un chip Xilinx Zynq donde se encuentran un microprocesador y una FPGA que se comunican por un bus de datos en el interior del chip. El microprocesador que funciona en Linux tiempo real descarga en la FPGA el código que se va a ejecutar y realiza la comunicación E/S del chip a través del WiFi. La FPGA lee todos los canales en paralelo y de manera síncrona. Como el número de entradas digitales con las que cuenta la myRIO es menor que el número de micrófonos se hace que cada canal sea capaz de recoger la información de dos sensores (uno en el flanco de subida del reloj de 2 MHz y otro en el flanco de bajada). De este modo, con 32 entradas digitales, se captura la salida de 64 MEMS.

La myRIO cuenta con un conversor D/A a través del cual se genera el pulso de transmisión. Una vez generado, la myRIO se comunica con un amplificador externo de audio (Figura 18). El amplificador se conecta al tweeter ajustando la impedancia y generando una potencia de audio significativa para transmitir la señal.



Figura 18: Amplificador externo

Por otra parte, la myRIO se conecta físicamente al array acústico (*Figura 19*) para recoger las señales capturadas por cada micrófono y vía WiFi a un PC que se utiliza como interfaz de usuario y unidad de procesamiento.



*Figura 19 myRIO conectada al array acústico MEMS [30]*

### 3.2 CONFORMADO Y PREPROCESADO

El hecho de utilizar un array planar 8x8 hace que el número de canales requeridos sea de  $N^2 = 64$ . Hay que tener en cuenta que la complejidad y los costes están directamente relacionados con el número de canales/sensores del sistema [30].

El software encargado de conformar y preprocesar las señales acústicas ha sido desarrollado en LabVIEW íntegramente por el equipo del GPA (Grupo de Procesado en Array) de la universidad de Valladolid.

El software en cuestión se denomina SAM (Sistema Acústico Mems), y cuenta con diversos programas:

- CaSAM: Captura y guarda a disco las señales temporales permitiendo ver las imágenes acústicas pero sin guardarlas a disco.
- ViSAM: Permite visualizar las señales temporales y las imágenes acústicas pero sin guardarlas a disco.
- BaSAM: Realiza un procesamiento por lotes (*bacht*) posterior a la adquisición de las señales, calculando las imágenes acústicas mediante conformación y permitiendo guardar a disco las imágenes obtenidas.

Para conformar las señales en banda estrecha, se utiliza una técnica de *beamforming* que separa las componentes en fase y cuadratura de las señales y las multiplica por un único

desfase. Además, al conformar, se aplica un enventanado espacial a partir de la ganancia de cada uno de los sensores.

Al utilizar una frecuencia única, se podría aplicar un algoritmo *delay-sum* que retarda las señales (por cada haz se aplica un retardo distinto), mientras que, en caso de utilizar múltiples tonos, se utiliza un algoritmo basado en la transformada de Fourier rápida FFT (*Fast Fourier Transform*). *Delay-sum* necesita mucha memoria para almacenar las señales retardadas mientras que el algoritmo basado en la FFT sólo utiliza desfase siendo por ello óptimo cuando se envían múltiples tonos en la señal.

Por último, los algoritmos de conformado asumen el modelo de onda plana para la propagación de las señales, el aire como un medio homogéneo y que los micrófonos tienen todos diagrama de radiación omnidireccional y misma respuesta frecuencial.

La frecuencia y ancho de pulso de la onda acústica transmitida por el sistema condicionan su funcionamiento por lo que se deben escoger de manera inteligente.

La etapa de preprocesado de las imágenes (antes de aplicar cualquier tipo de modelo o algoritmo de extracción de parámetros), ha sido necesaria para resolver algunos problemas que surgen al capturar perfiles acústicos.

Como se comenta en el apartado “Discusión sobre las reflexiones producidas”, los micro movimientos que el individuo escaneado realiza mientras se captura su perfil acústico afectan notablemente a las reflexiones producidas y por ende a la señal recibida en los micrófonos.

Motivado por este problema de los micro movimientos, se decide emplear una estrategia de promediado que se aplica sobre las imágenes ya conformadas mejorando notablemente los resultados. No obstante, hay que tener en cuenta que un promedio de  $M$  imágenes incrementa el tiempo de adquisición para cada muestra de un perfil acústico en un factor  $M$ .

La otra estrategia de preprocesado utilizada, se ha aplicado antes del conformado y consiste en eliminar el *clutter* de las señales temporales capturadas. El término *clutter* se refiere, en el ámbito RADAR a los ecos no deseados que forman parte del escenario y son completamente independientes del individuo objeto de escáner. Para eliminar este ruido indeseado, se captura previamente el escenario de adquisición sin el individuo para eliminarlo posteriormente de las capturas realizadas con el individuo. Esta estrategia, aunque apenas necesaria al utilizar una cámara anecoica como escenario [30], permite mejorar la calidad de las imágenes acústicas obtenidas eliminando ecos ajenos al individuo.

### **3.2.1 ANCHO DE PULSO**

La elección del valor del ancho de pulso supone una relación de compromiso. Es decir, cuanto mayor sea el ancho de pulso, mayor será la energía recibida pero peor será la resolución en rango y peor será el efecto del pulso directo por lo que no es una decisión trivial [6].

Típicamente, en los sistemas RADAR clásicos, existe un rango mínimo a partir del cual el sistema puede medir. Esto es así debido al pulso directo que reciben los micrófonos

directamente de la señal acústica generada en el tweeter, es decir, sin llegar a reflejarse en ningún blanco. Esto ocurre en los valores de rangos cercanos al array y como es lógico, si el ancho del pulso es mayor, este efecto se extenderá más en rango y con mayor energía.

Por otra parte, el ancho de pulso condiciona la resolución en rango. Si la separación en rango entre dos objetos es comparable al ancho de pulso, entonces será imposible distinguir los dos blancos. Por tanto, la resolución en rango es inversamente proporcional al ancho de pulso.

En contraposición, si el ancho de pulso es mayor, la onda acústica contiene más energía y el array será capaz de detectar blancos más débiles por lo que aumenta la sensibilidad del sistema.

Para este trabajo, tras probar con tiempos de pulso de  $1ms$  y  $2ms$  se ha decidido utilizar  $1ms$  ya que se prioriza no perder resolución en rango para diferenciar correctamente los elementos de reflexión del individuo que se encuentran relativamente cercanos en rango.

### 3.2.2 FRECUENCIA DE LA SEÑAL ACÚSTICA

La elección de la frecuencia de nuevo supone otra relación de compromiso. Las frecuencias demasiado altas cometen aliasing espacial. Si esto ocurre, se producen *grating lobes* y esto es un efecto indeseado al ser confundibles con el lóbulo principal. Por otra parte, la frecuencia es inversamente proporcional al ancho del lóbulo principal y este determina la directividad del array por lo que un ancho de lóbulo menor será capaz de distinguir con mayor precisión blancos cercanos.

La manera de poder observar la dependencia de estos efectos con los parámetros del sistema es observando su diagrama de radiación para distintos ángulos de apuntamiento. El diagrama de radiación del array permite visualizar la respuesta espacial del mismo y como se explicará con mayor detalle en el capítulo 4, se puede construir de manera teórica para diseñar de manera inteligente el sistema.

Es importante también tener en cuenta la región de interés puesto que será más difícil encontrar *grating lobes* en un espacio angular menor. En el caso de estudio, como se justificará más adelante, se ha escogido un área de barrido tanto en azimut como en elevación de  $-30^\circ$  a  $30^\circ$ .

Dado que el array planar es cuadrado y cuenta con  $8 \times 8$  sensores equiespaciados, para estudiar la frecuencia óptima será suficiente con observar cómo responde un array lineal de 8 sensores espaciados.

Primeramente, se ilustra para un ángulo de apuntamiento (*steering angle*)  $\vartheta = 0$ , como las mayores frecuencias consiguen un array espacialmente más selectivo, es decir con un ancho de lóbulo principal medido a  $3db$  menor (*Figuras 20-24*).

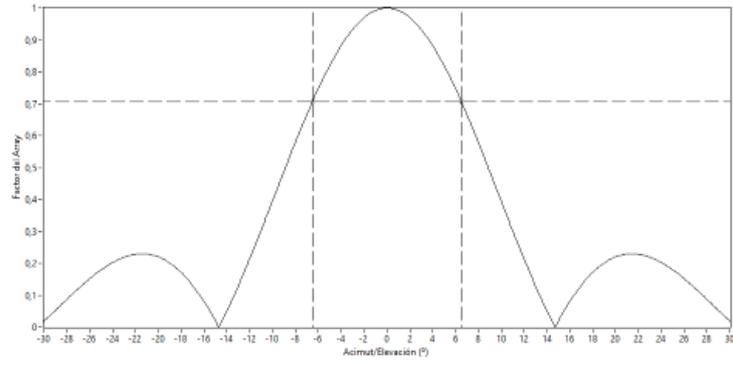


Figura 20: Diagrama de radiación de un array lineal con  $f = 8$  KHz

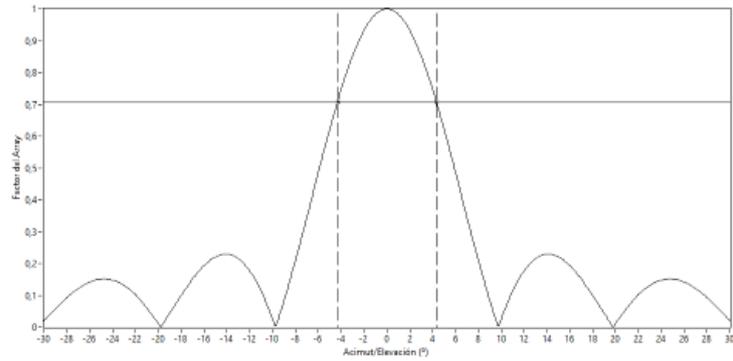


Figura 21: Diagrama de radiación de un array lineal con  $f = 12$  KHz

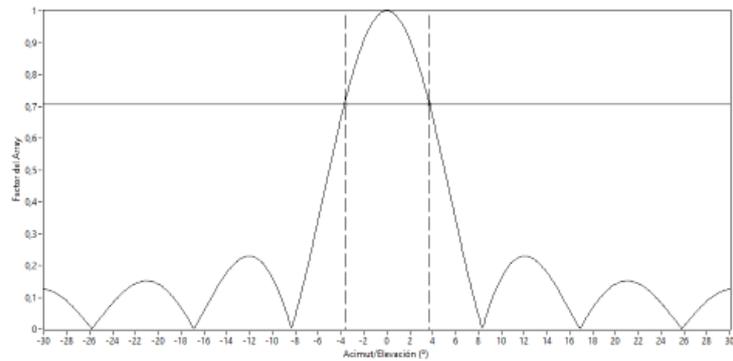


Figura 22: Diagrama de radiación de un array lineal con  $f = 14$  KHz

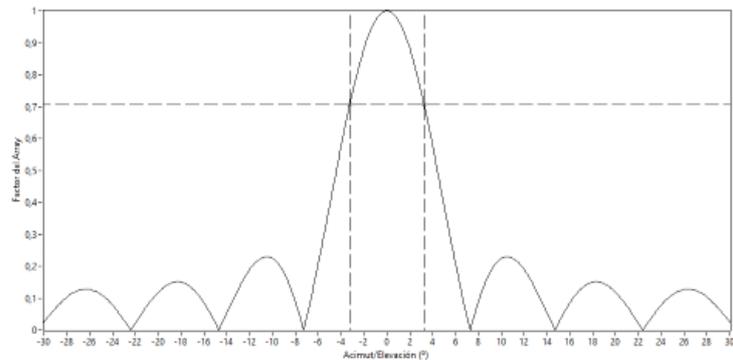


Figura 23: Diagrama de radiación de un array lineal con  $f = 16$  KHz

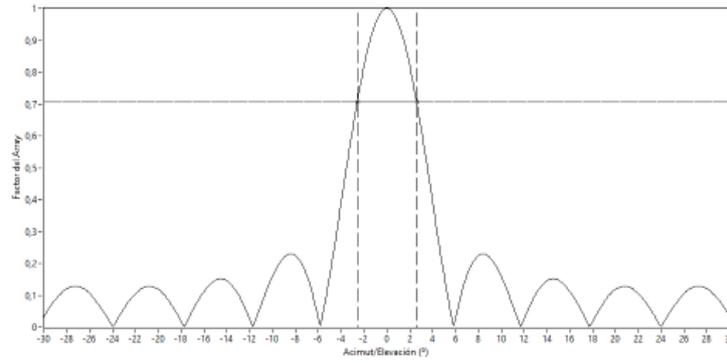


Figura 24: Diagrama de radiación de un array lineal con  $f = 20 \text{ KHz}$

Entonces, aquella frecuencia que teniendo la mejor resolución espacial garantice no cometer aliasing en el rango de  $-30^\circ$  a  $30^\circ$  será la mejor. Para ello, basta con observar los diagramas correspondientes a distintas frecuencias para un ángulo de apuntamiento en el límite del rango de ángulos a considerar, es decir  $30^\circ$  (Figura 25).

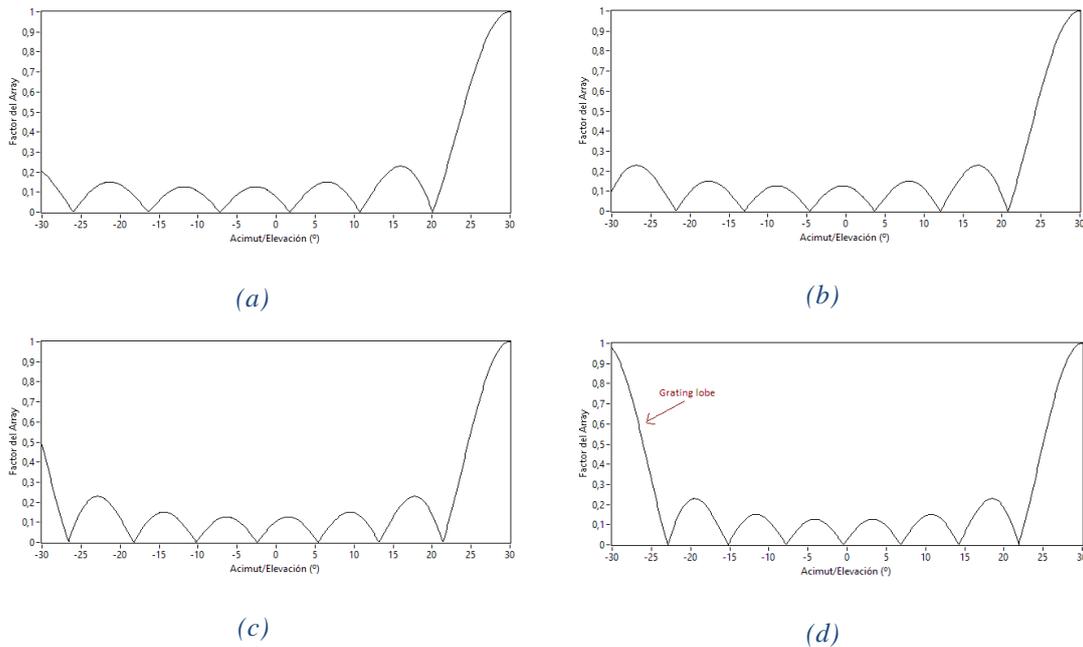


Figura 25: Diagramas de radiación de un array lineal para  $\theta = 30^\circ$  a distintas frecuencias: (a) 13 KHz, (b) 14 KHz, (c) 15 KHz, (d) 16 KHz

Teniendo en cuenta que utilizando las frecuencias de 15 y 16 KHz, un eco proveniente de  $\pm 30^\circ$  provocaría un efecto de aliasing espacial lo suficientemente notable en las imágenes, se ha elegido utilizar una frecuencia de 14 KHz que sin producir *grating lobes* promete mejor resolución.

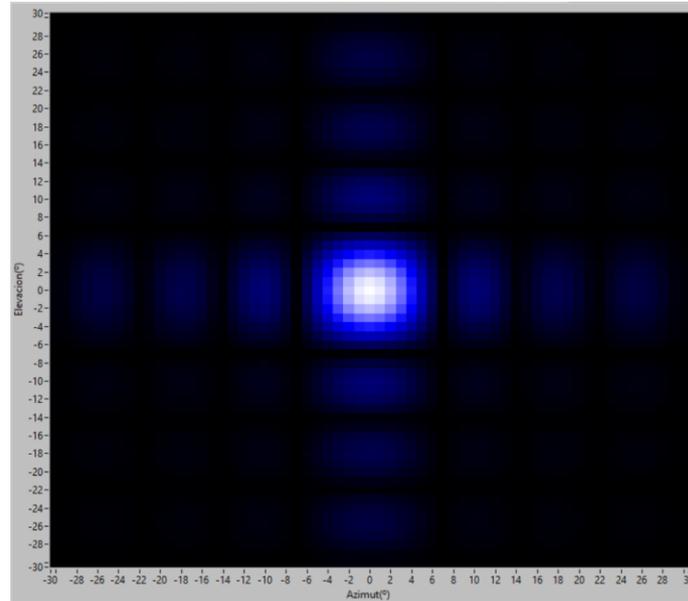
### 3.3 CONSIDERACIONES DE ADQUISICIÓN

#### 3.3.1 ESCENARIO

Para simplificar el escenario de procesado y tener en cuenta únicamente las reflexiones acústicas producidas por el individuo “capturado”, el sistema de adquisición se encuentra

dentro de una cámara anecoica. Esto permite absorber casi la totalidad de las reflexiones producidas por las ondas acústicas en las paredes, suelo y techo de la sala. Para reducir las reflexiones acústicas, la cámara anecoica se ha construido a partir de cuñas piramidales hechas de espumas cuyo coeficiente de absorción acústica (directamente proporcional a la absorción del material) es elevado.

Otro tema importante es la posición física del sistema de adquisición. Al tratarse de un array planar cuadrado con sensores equiespaciados en ambas direcciones, se obtiene un diagrama de radiación simétrico (*figura 26*). Es por ello que el sistema proporciona la misma resolución en ambas coordenadas vertical y horizontal.

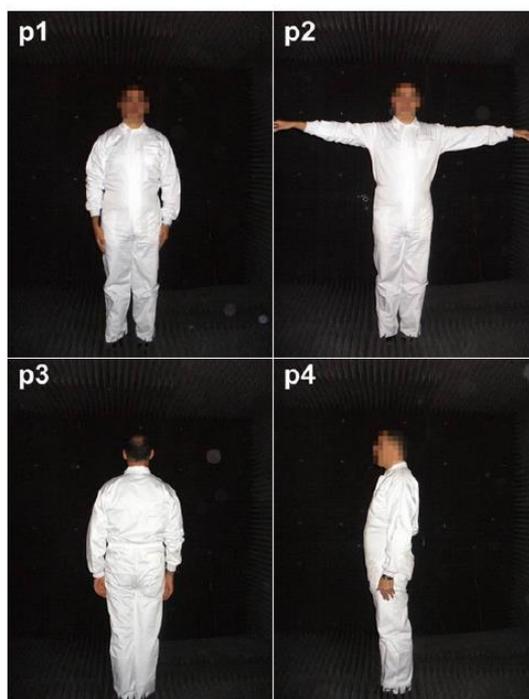


*Figura 26: Diagrama de radiación del array a 14Khz*

Por otra parte, la altura respecto al suelo del centro del array si es notablemente significativa, ya que repercute directamente en el espacio capturado en la imagen. Lo ideal sería que este se encuentre en el centro entre el suelo y la altura del individuo a capturar, para que la persona se encuentre centrada en elevación. Siguiendo este razonamiento colocar el array a una distancia de aproximadamente 90 centímetros del suelo es una buena decisión con el fin de centrar al sujeto promedio en la captura.

### **3.3.2 POSICIÓN DEL INDIVIDUO Y LÍMITES ESPACIALES**

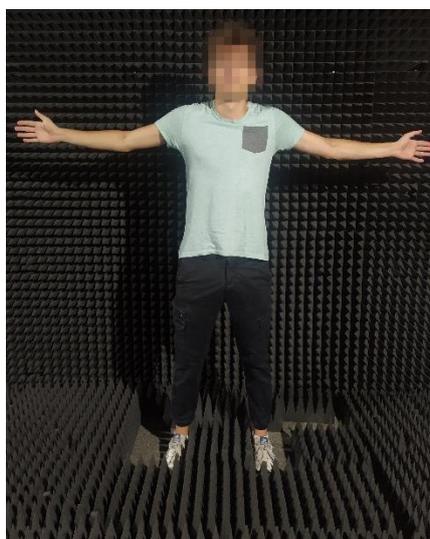
La captura de los rasgos antropométricos más significativos depende notablemente de la postura del individuo a la hora de ser fotografiado por el sistema acústico. Por ejemplo, la cantidad de información antropométrica obtenida si el individuo extiende los brazos no es comparable a la que se obtiene si los esconde detrás de la espalda. En estudios anteriores [6], se consideraron cuatro posibles posiciones (*Figura 27*) sobre las que se ha partido en esta ocasión.



*Figura 27: Posiciones para adquisición planteadas en anteriores estudios [6]*

Para el trabajo a desarrollar, en vez de utilizar las 4 posiciones, se ha utilizado una sola que proporcione suficiente información biométrica con el fin de reducir el tiempo de adquisición de las imágenes.

Partiendo de la posición (*p2*), se han aplicado algunos cambios hasta llegar a la que ha sido tomada como referencia para las capturas tomadas finalmente en el proyecto (*figura 28*).



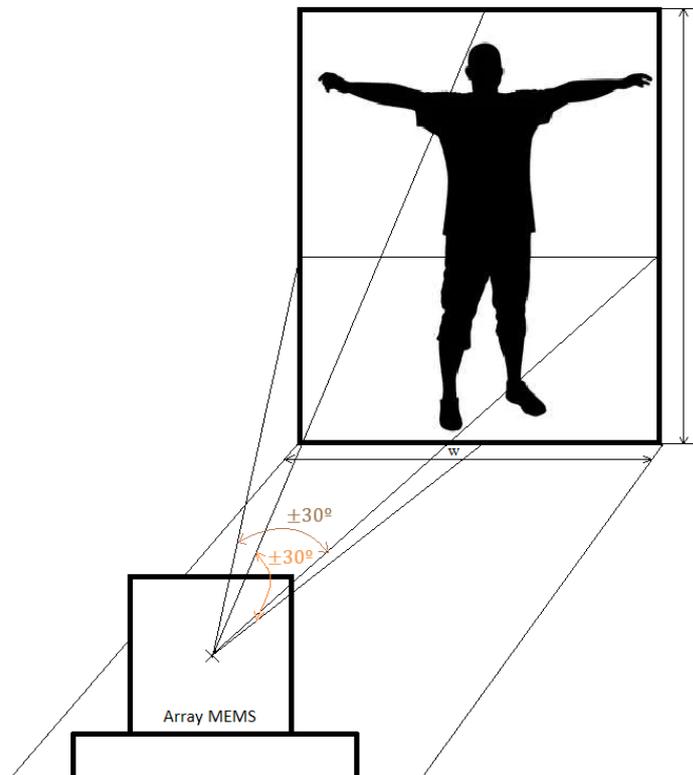
*Figura 28: Posición del individuo escogida para las capturas*

Como resumen, se ha decidido separar las piernas en lugar de mantener los pies juntos con el objetivo de mejorar la detección de los ecos acústicos asociados a cada una de las piernas. También se toman las nuevas imágenes con las manos extendidas en lugar del

puño lo que incrementa su detección al ser mayor la superficie sobre la que incide la señal acústica.

Respecto a la distancia al array, se ha colocado una marca en el suelo a aproximadamente 2 metros del array para que todos los individuos coloquen los pies a la misma distancia.

Es importante determinar también unos rangos de captura para no aumentar el volumen de información innecesariamente. Situando una persona a unos 2 metros del array y teniendo en cuenta la altura a la que está situado el centro del sistema (90 centímetros), un rango de ángulos de  $\pm 30^\circ$  debería ser más que suficiente para realizar las capturas que serán objeto de estudio. Aplicando la conversión correspondiente entre las coordenadas polares que maneja el array y las coordenadas cartesianas, se tiene que el sistema sería capaz de capturar, en horizontal, desde el suelo hasta unos 2.06 metros y en vertical algo menos de 1.2 metros desde el centro del array a cada lado, es decir observando la imagen inferior (*Figura 29*), se tendría que  $h \approx 2.06 \text{ m}$  y  $w \approx 2.4 \text{ m}$ .



*Figura 29: Disposición del individuo frente al array acústico*

Esto, acorde con algunos registros como “Alcance máximo horizontal” y “Estatura” recogidos en estudios antropométricos de la población laboral española [20], puede dar una idea del desempeño del sistema. Los valores seleccionados permitirían capturar todas las reflexiones acústicas producidas por casi la totalidad de los posibles individuos que utilicen el sistema.

Por último, otro factor que se debe tener en cuenta es la resolución del muestreo en acimut y elevación a la hora de tratar con la imagen acústica final como una matriz de información. Si se toma un ángulo menor entre haces de la imagen acústica, entonces el volumen de información crecerá notablemente. Contar con más información, siempre que

sea significativa, resultará útil, pero esto conlleva un aumento notable en la carga computacional por lo que escoger un valor muy pequeño no es una buena solución. Una buena relación de compromiso se ha encontrado utilizando  $1^\circ$  entre haces. En cartesianas, a una distancia de 2 metros del array, esto supone contar con una muestra cada 3.5cm por lo que parece suficiente para recoger las reflexiones acústicas producidas por el cuerpo humano. Aunque se podría haber escogido un número menor, el valor de  $1^\circ$  permite contar en cada plano acimut-elevación con 3721 ( $61^2$ ) muestras lo cual no es un valor excesivamente grande.

Respecto al muestreo y margen de captura en rango, dado que la persona se encuentra centrada a una distancia de 2 metros, se han capturado las imágenes en el rango de 150cm a 250cm. Siguiendo la misma filosofía que para la resolución angular, se han tomado muestras cada centímetro por lo que se espera del sistema una mayor resolución en rango.

### ***3.3.3 DISCUSIÓN SOBRE LAS REFLEXIONES ACÚSTICAS***

Para entender cómo son las imágenes acústicas conseguidas, es importante estudiar las propiedades de propagación de las señales acústicas (reflexión, difracción, dispersión y absorción) cuando se encuentran con cuerpos humanos [34].

La intensidad de los ecos reflejados depende de la forma, textura y características acústicas de la superficie de reflexión además de su orientación relativa al array. Debido a la compleja superficie del cuerpo humano, la señal se refleja de forma difusa en el conjunto [8].

En estudios orientados a la detección de humanos frente a otros objetos [8], se ha considerado que las reflexiones acústicas se centran mayoritariamente en el torso y la cabeza. Además, dependiendo de la distancia y ubicación del ser humano, si la resolución del conformador no es suficientemente buena, múltiples ecos pueden mezclarse en uno solo o incluso reflejar varias partes del torso de manera indistinta. Sin embargo, en el ámbito de la identificación, las reflexiones de torso y cabeza no son suficientes para recoger la suficiente información biométrica de la persona por lo que se deben caracterizar más elementos de reflexión, aunque no aparezcan tan claramente en las imágenes.

Por otra parte las ondas acústicas se atenúan al viajar por un medio. Esto quiere decir que los ecos asociados a los blancos más alejados del array se recibirán con menor intensidad y por ende será más difícil su detección. Además, esta atenuación es directamente proporcional a la frecuencia por lo que el uso de frecuencias más altas cubre distancias menores [16].

El producto de la velocidad del sonido dentro de cada medio y su densidad determinan el porcentaje de haz reflejado o transmitido al pasar de un medio a otro. Este valor recibe el nombre de impedancia acústica y recoge características de reflexión o transmisión de un medio. De esta manera, al alcanzar una onda acústica una superficie de discontinuidad entre dos medios, la amplitud del eco de retorno es proporcional a la diferencia de impedancia acústica entre dos tejidos cuando la onda acústica atraviesa la interfaz por lo que mayores diferencias producirán reflexiones más fuertes [16].

La tabla a continuación (*Figura 30*), recoge los valores de impedancia acústica de algunos tejidos humanos y del aire. Debido a la baja impedancia acústica del aire, cuando la onda se encuentra con una interfaz de aire–tejido blando, casi todo el sonido se refleja [16].

Tissue or Substance	Acoustic Impedance*
Air	0.0004
Fat	1.38
Water (50°C)	1.54
Brain	1.58
Blood	1.61
Kidney	1.62
Liver	1.65
Muscle	1.70
Lens	1.84
Bone	7.8

\*Acoustic impedance (Z) =  $\times 10^6$  kg/m<sup>2</sup>sec.  
 Data from Curry TS III, Dowdey JE, Murry RC Jr: Christensen's  
 Physics of Diagnostic Radiology, 4th ed. Philadelphia: Lea & Febiger,  
 1990.

*Figura 30: Impedancias acústicas de algunos tejidos humanos* [16]

Pero no toda la energía apunta directamente a nuestro cuerpo, de hecho, la mayoría de las ondas reflejan sobre nuestra ropa por lo que las propiedades acústicas de los tejidos también deben ser tenidas en cuenta. Por ejemplo, el lino o el poliéster son tejidos que cuentan con buenas propiedades de absorción acústica [35] por lo que empeorarán la calidad del sistema biométrico absorbiendo gran parte de la energía.

Además, la intensidad de los ecos reflejados depende de la superficie de reflexión como se ha comentado al principio. Esto hace que surja una dificultad importante con la que tendrá que lidiar el sistema biométrico. Básicamente, el hecho de llevar una prenda de ropa de una holgura, textura o material diferentes causará notables diferencias sobre los ecos acústicos obtenidos. De hecho, llevar un pantalón corto o largo influye en cierta manera en las capturas acústicas.

El ángulo de reflexión dependerá también del ángulo de incidencia por lo que es entendible que dependiendo de la forma de la superficie sobre la que incide, las reflexiones producidas sean distintas.

Es inevitable que cuando una persona está siendo capturada realice micro movimientos que la hagan cambiar ligeramente la postura u orientación de ciertas partes del cuerpo. Dependiendo de la postura y la orientación, algunas zonas de reflexión pueden quedar ortogonales a la línea recta que los une con el array o no y esto los hace ser más o menos reflejados [10].

Por último, cabe realizar una pequeña discusión sobre la reflexión difusa y especular. Este trabajo asume que cuando se radia una onda acústica, cualquier parte del cuerpo devuelve un eco hacia la posición del array, esto es reflexión difusa. Es decir, se refleja una onda acústica que viaja en todas direcciones. Si el cuerpo radiado produjese una reflexión especular perfecta, el array solo “vería” ese blanco si la reflexión se recibe en el plano del

array. Esto depende en parte de la superficie sobre la que incide la onda. De hecho, una esfera típicamente produce una reflexión especular a la misma dirección sobre la que incide la onda. Si una persona estuviese formada por blancos puntuales, se podría asumir que siempre se recogería energía de vuelta en la posición del array.

# 4

## Capítulo 4. DEFINICIÓN DEL MODELO

---

A la hora de elaborar un modelo hay que tener en cuenta varias consideraciones. George E. P. Box dijo que todos los modelos son erróneos, pero algunos son útiles: “All models are wrong, but some are useful” (George E. P. Box, 1978). Esta cita no se debe de perder de vista a la hora de elaborar cualquier tipo de modelo, y es que un modelo representa una abstracción de la realidad y por eso nunca refleja de manera exacta lo que se estudia.

Se pretende generar a partir de las capturas acústicas, un modelo que sintetice el gran volumen de información que contienen las imágenes en un conjunto de parámetros que sean suficientemente representativos del individuo. Este modelo nunca va a representar a través de sus parámetros a la imagen acústica con exactitud, pero puede ser una buena simplificación.

El primer razonamiento que se ha llevado a cabo en el proyecto ha sido explotar la información teórica conocida sobre la respuesta espacial del array acústico. La idea que reside detrás de este razonamiento es la siguiente: Si se sabe cuál es la respuesta teórica del array ante un blanco acústico puntual, se puede calcular cuál será la respuesta del array ante un número finito de blancos.

Para comenzar, se ha planteado a través de una aproximación sucesiva un primer modelo que suma de manera independiente la respuesta en módulo del array ante cada blanco. La denominada “aproximación sucesiva” se entiende como una manera secuencial de abordar el problema aumentando por fases la dimensionalidad de este.

Este modelo, aunque bastante simplificado no tiene en cuenta el carácter de las interferencias (constructivas o destructivas) entre los distintos blancos por lo que surge la idea de un nuevo modelo más riguroso que sí tiene en cuenta las fases y no sólo los módulos.

Por último, se define un nuevo modelo basado en los máximos 3D presentes en la captura que busca caracterizar de manera correcta a los individuos escaneados a costa de una peor representación de la imagen.

### 4.1 APROXIMACIÓN SUCESIVA. MODELO REAL

¿Por qué utilizar una aproximación sucesiva? Entender los detalles del problema es fundamental para poder elaborar un buen modelo cuya extracción de parámetros sea sencilla. Esto permite entender primero un problema más sencillo que sirve de base para resolver el siguiente. Así pues, en la primera aproximación 1D que se realiza sólo se tiene

en cuenta el acimut como variable espacial para después ir añadiendo a la ecuación la variable elevación y por último la variable rango.

Este modelo asume que los individuos están formados por un número finito de elementos de reflexión y por ende sus respuestas acústicas se pueden modelar como un número finito de blancos. Además, también asume que los blancos son independientes y no existe interacción alguna entre sí lo que permite construir una imagen acústica simplemente como el agregado del número de blancos presentes en la imagen.

#### 4.1.1 SÍNTESIS 1D

Como primera aproximación, se supone un array lineal a través del cual se captura un rango concreto. Es decir, la única variable independiente es el acimut.

Se supone como primera simplificación, que la respuesta acústica de una persona al array lineal está formada por 3 únicos blancos: torso, brazo izquierdo y brazo derecho por lo que acorde con la hipótesis planteada, se recibirán en el array 3 ecos de distinta amplitud procedentes de ángulos diferentes. (Tabla 1)

Blancos de la persona:	Ganancia $k$ :	Acimut $\theta$ :
Torso	1	0°
Brazo izquierdo	0.5	10°
Brazo Derecho	0.5	-10°

Tabla 1: Modelo simplificado de 3 blancos 1D para una persona

El eco acústico recibido de cada uno de estos blancos queda caracterizado por el diagrama de radiación correspondiente a la respuesta espacial del array. Por lo que, en un conformador determinístico, el módulo de la señal recogida en el array lineal ante un blanco de ganancia  $k_i$  y acimut  $\theta_i$  es conocida. (Ecuación 4.1)

$$|F_i(\theta)| = k_i \cdot \left| \frac{\sin \left[ \frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta) \right]}{\sin \left[ \frac{\omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta) \right]} \right| \quad (4.1)$$

Donde  $\omega$  es la pulsación de trabajo,  $N$  el número de sensores,  $c$  la velocidad de propagación del sonido y  $d$  la separación entre sensores. La variable independiente  $\theta$ , corresponde al ángulo de apuntamiento del array. Como se presentó en el capítulo 3:

- $N = 8$  sensores
- $\omega = 2\pi f = 2\pi \cdot 14 \cdot 10^3 \text{ rad/s}$
- $d = 2.5\text{cm}$

Respecto a la velocidad de propagación del sonido, su valor depende exclusivamente del medio por el que se propaga la onda. En el aire, muchas veces se utiliza  $c = 340 \text{ m/s}$  como aproximación, pero lo cierto es que este valor depende de varios factores. Acorde con la hipótesis adiabática de Laplace y la ecuación de gas ideal que resulta de la ley de Boyle, para un gas ideal, la velocidad del sonido viene dada por:

$$c = [RT(\gamma/M)]^{1/2} \quad (4.2)$$

Donde  $\gamma$ ,  $M$ ,  $R$  y  $T$  corresponden a la constante adiabática, masa molar, constante universal de los gases ideales y temperatura absoluta, respectivamente [36], [37]. En resumen, la velocidad del sonido depende de factores como la humedad o la presión, pero más aún de la temperatura. De hecho, de la ecuación anterior (*Ecuación 4.2*), se puede deducir como la velocidad del sonido aumenta unos  $0.6m/s$  por cada  $1^\circ C$  que aumenta la temperatura. Como referencia, en aire seco estándar a  $0^\circ C$  y una presión barométrica de  $101.325kPa$ , se tiene que  $c = 331.29 m/s$  [36]. No obstante, el software encargado de generar las imágenes acústicas tiene en cuenta en el proceso de *beamforming* una velocidad constante de  $c = 345.4 m/s$  por lo que se elige este valor para el modelo a sabiendas de que pequeños cambios en la temperatura afectan notablemente a la velocidad del sonido.

A la hora de programar este modelo sintético, es interesante destacar que la ecuación (*Ecuación 4.1*) presenta una indeterminación en  $\theta = \theta_i$  que se debe resolver para obtener de manera programática  $|F_i(\theta)|$ . Sencillamente, se puede aplicar la regla de l'Hôpital sobre la ecuación (*Ecuación 4.1*) con lo que se llega a (*Ecuación 4.3*).

$$\lim_{\theta \rightarrow \theta_i} F_i(\theta) = \lim_{\theta \rightarrow \theta_i} \left( \frac{\cos\left(\frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta)\right) \cdot \left(-\frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot \cos(\theta)\right)}{\cos\left(\frac{\omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta)\right) \cdot \left(-\frac{\omega}{2} \cdot \frac{d}{c} \cdot \cos(\theta)\right)} \right) \quad (4.3)$$

En este momento, resolver el límite de la ecuación (*Ecuación 4.3*) no supone ningún problema. Así, se obtiene que  $|F_i(\theta_i)| = N$

$$\lim_{\theta \rightarrow \theta_i} F_i(\theta) = \frac{\cos\left(\frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot 0\right) \cdot N}{\cos\left(\frac{\omega}{2} \cdot \frac{d}{c} \cdot 0\right)} = N \quad (4.4)$$

En base a esto, se puede generar la respuesta espacial sintetizada para los 3 blancos correspondientes como una combinación lineal de las respuestas a cada uno de ellos (*Ecuación 4.5*).

$$P_{Persona}(\theta) = |F_1(\theta)| + |F_2(\theta)| + |F_3(\theta)| \quad (4.5)$$

Gráficamente, se puede ver a continuación (*Figura 31*), la respuesta individual del array a cada uno de los blancos y en azul, la señal sintetizada de la persona compuesta por el modelo simplificado de 3 blancos considerado. No hay que olvidar que según se expuso en el capítulo 3, la resolución del sistema es de  $1^\circ$ .

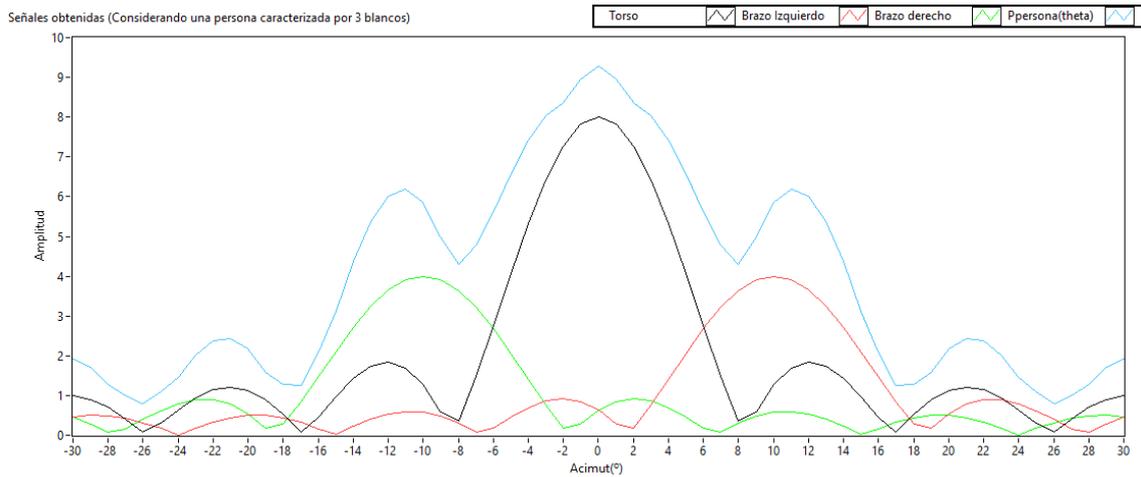


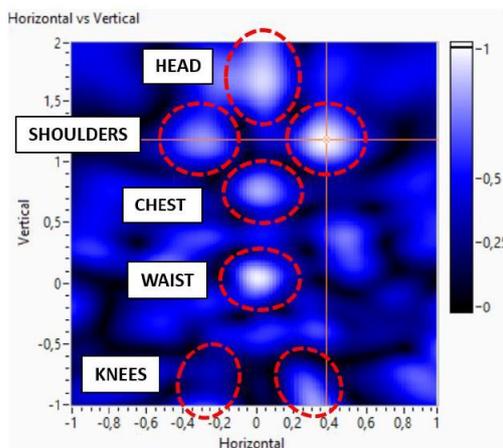
Figura 31: Señales  $|F_1(\theta)|$ ,  $|F_2(\theta)|$ ,  $|F_3(\theta)|$  y  $P_{Person}(\theta)$

En esta imagen se puede apreciar también cómo los máximos correspondientes al lóbulo principal de los brazos izquierdo y derecho son desplazados ligeramente a causa de los lóbulos secundarios del torso que es el blanco más fuerte. Esto anticipa un problema con los lóbulos secundarios de los blancos más fuertes.

#### 4.1.2 SÍNTESIS 2D

Se incrementa a continuación la dimensionalidad del problema agregando la elevación como variable independiente y planteando así un modelo más elaborado de la respuesta acústica de una persona. Se puede interpretar esta síntesis en 2 dimensiones como cortes en el plano acimut-elevación de una imagen acústica tomada con el array planar presentado en el capítulo 3.

Imágenes capturadas en estudios anteriores de un maniquí a 20 KHz (Figura 32) motivan a utilizar cabeza, hombros, pecho, cintura y rodillas como simplificación de los blancos de una persona.



(a)

(b)

Figura 32: (a) Imagen acústica tomada de un maniquí, (b) Imagen óptica del maniquí capturado

Esto es un modelo aún muy simplificado y de hecho hay que tener en cuenta que la imagen del maniquí fue tomada a partir de las rodillas y que sus brazos no se encuentran extendidos como se plantea en el capítulo anterior a la hora de tomar las imágenes.

De este modo, se considerará ahora que la respuesta acústica de una persona está formada por los siguientes 7 blancos con distinta amplitud y procedentes de ángulos diferentes (*Tabla 2*):

Blancos de la persona:	Ganancia $k$ :	Acimut $\theta$ :	Elevación $\varphi$ :
Cabeza	0,8	0°	18°
Pecho	0,3	0°	8°
Hombro Izquierdo	0,4	7°	14°
Hombro Derecho	0,4	-7°	14°
Cintura	1	0°	-1°
Rodilla Derecha	0,3	-4°	-14°
Rodilla Izquierda	0,3	4°	-14°

*Tabla 2: Modelo simplificado de 7 blancos 2D para una persona*

Así, siguiendo el mismo razonamiento que en el apartado anterior, el módulo de la señal recibida en el array planar ante un blanco puntual de ganancia  $k_i$ , acimut  $\theta_i$  y elevación  $\varphi_i$  viene dado por el producto de los diagramas de radiación 1D (*Ecuación 4.1*) en ambas coordenadas tal cual se muestra en la siguiente ecuación:

$$|F_i(\theta, \varphi)| = k_i \cdot \left| \frac{\sin \left[ \frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta) \right]}{\sin \left[ \frac{\omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta) \right]} \right| \cdot \left| \frac{\sin \left[ \frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot (\sin \varphi_i - \sin \varphi) \right]}{\sin \left[ \frac{\omega}{2} \cdot \frac{d}{c} \cdot (\sin \varphi_i - \sin \varphi) \right]} \right| \quad (4.6)$$

Y de nuevo al igual que en la (*Ecuación 4.5*), se puede caracterizar la respuesta de la persona como la suma de la respuesta ante cada blanco:

$$P_{Persona}(\theta, \varphi) = \sum_{i=1}^7 |F_i(\theta, \varphi)| \quad (4.7)$$

Gráficamente se obtiene una imagen normalizada como la siguiente (*Figura 33*), donde se ha marcado con un cursor rojo la ubicación de cada blanco:

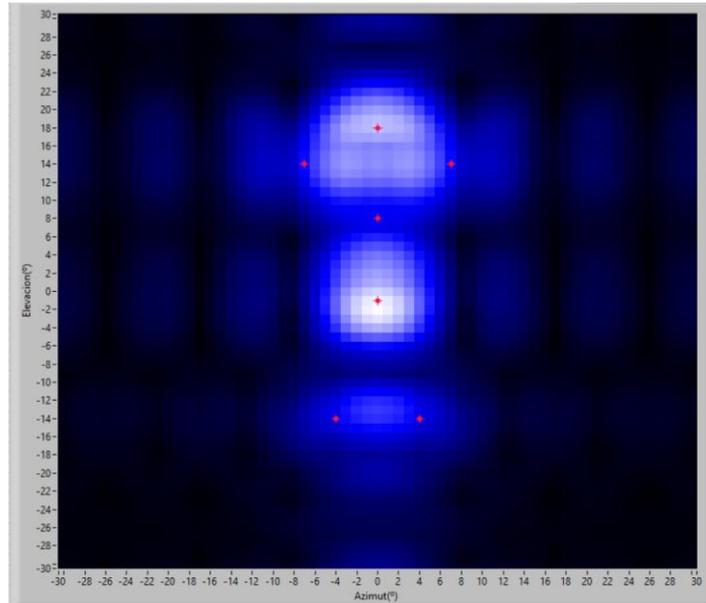


Figura 33: Imagen 2D  $P_{\text{persona}}(\theta, \varphi)$  generada a partir de 7 blancos sintéticos

En esta imagen se percibe otro problema que tiene que ver con la resolución del array. Si los blancos no se encuentran lo suficientemente espaciados (como en el caso sintético planteado), puede ser difícil diferenciarlos entre sí. De hecho, el eco acústico de las rodillas parece fundirse en un blanco único y el del pecho parece inapreciable. Esto tiene que ver con el ancho del lóbulo principal del array cuyo ancho de banda a 3db se puede medir a través de su diagrama de radiación (Figura 34).

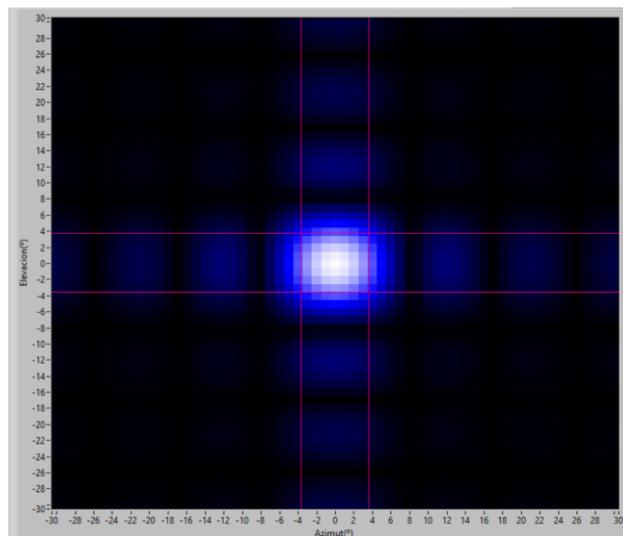


Figura 34: Ancho de banda a 3db del diagrama de radiación del array

### 4.1.3 SÍNTESIS 3D

El salto a las imágenes 3D implica hablar del pulso acústico transmitido. Se plantearon diversas alternativas para modelar dicho pulso (pulso triangular (Figura 35a), pulso triangular suavizado (Figura 35b), pulso gaussiano (Figura 36) ...). De todos ellos, a través del pulso gaussiano se obtuvieron los mejores resultados dado que el pulso acústico transmitido por el sistema de adquisición es también gaussiano.

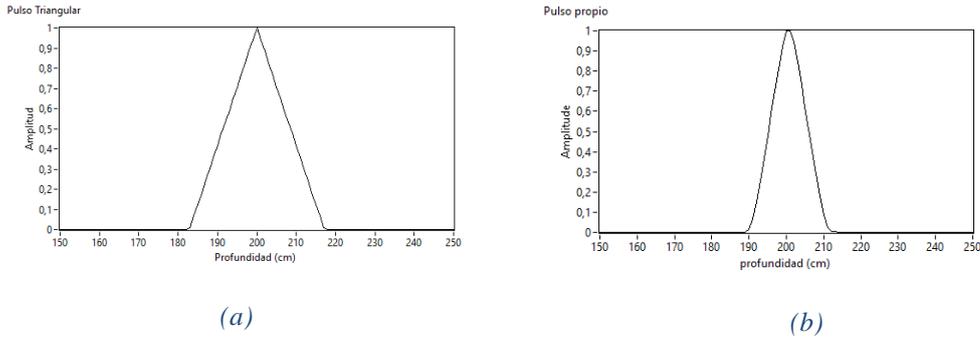


Figura 35: Primeras propuestas de pulsos utilizados para sintetizar las imágenes 3D

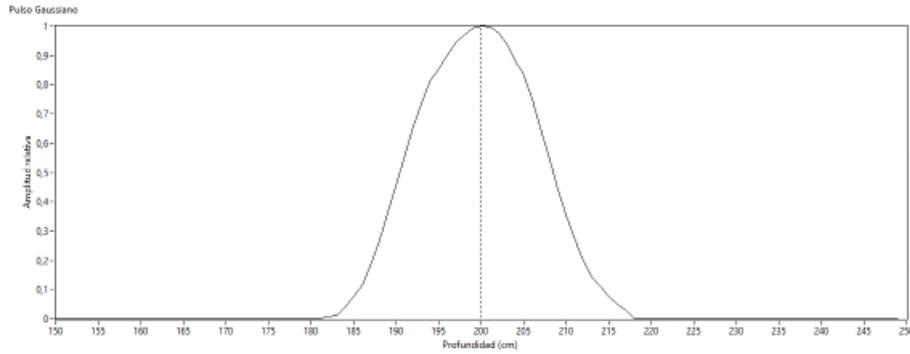


Figura 36: Pulso gaussiano de 1ms utilizado para sintetizar las imágenes 3D

De este modo y siguiendo con el razonamiento recurrente, el módulo de la señal recibida en el array planar ante un blanco puntual de ganancia  $k_i$ , acimut  $\theta_i$ , elevación  $\varphi_i$  y rango  $r_i$  viene dado por el producto del diagrama de radiación 2D calculado en el apartado anterior (Ecuación 4.6) con la función correspondiente al pulso gaussiano  $G_{pulse}(r)$

$$|F_i(\theta, \varphi, r)| = k_i \cdot \left| \frac{\sin \left[ \frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta) \right]}{\sin \left[ \frac{\omega}{2} \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta) \right]} \right| \cdot \left| \frac{\sin \left[ \frac{N \cdot \omega}{2} \cdot \frac{d}{c} \cdot (\sin \varphi_i - \sin \varphi) \right]}{\sin \left[ \frac{\omega}{2} \cdot \frac{d}{c} \cdot (\sin \varphi_i - \sin \varphi) \right]} \right| \cdot G_{pulse}(r, r_i) \quad (4.8)$$

De nuevo se considera el caso de estudio de los 7 blancos presentados en el apartado anterior, aunque se añade ahora en la siguiente tabla la información correspondiente acerca del rango (Tabla 3).

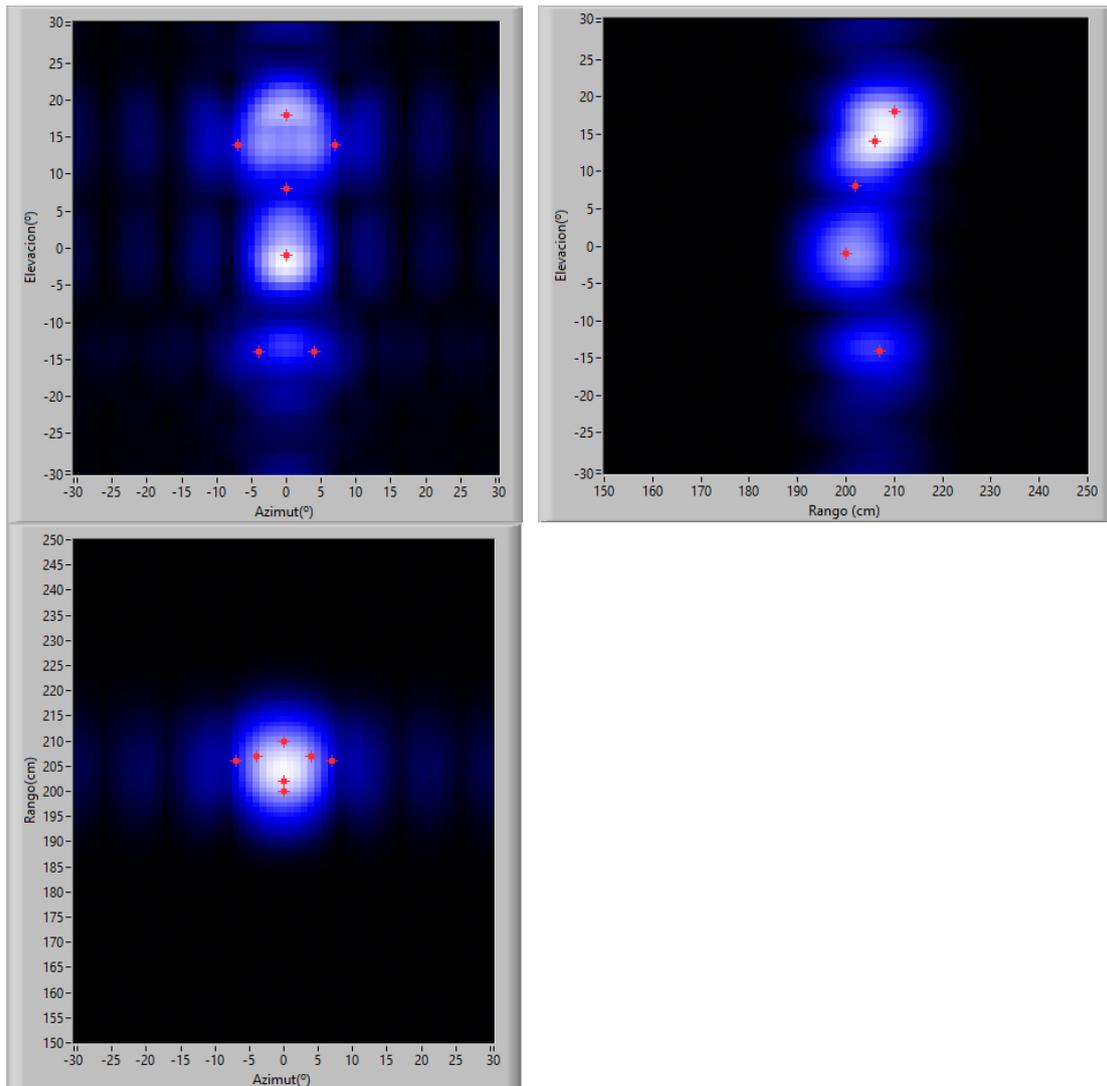
Blancos de la persona:	Ganancia $k$ :	Acimut $\theta$ :	Elevación $\varphi$ :	Rango $r$ :
Cabeza	0,8	0°	18°	210
Pecho	0.3	0°	8°	202
Hombro Izquierdo	0.4	7°	14°	206
Hombro Derecho	0,4	-7°	14°	206
Cintura	1	0°	-1°	200
Rodilla Derecha	0,3	-4°	-14°	207
Rodilla Izquierda	0,3	4°	-14°	207

Tabla 3: Modelo simplificado de 7 blancos 3D para una persona

En base a la información acerca de los blancos presentados, se puede construir la imagen acústica 3D sumando la respuesta individual del array a cada uno de los blancos según la siguiente ecuación:

$$P_{Persona}(\theta, \varphi, r) = \sum_{i=1}^7 |F_i(\theta, \varphi, r)| \quad (4.9)$$

Siguiendo este procedimiento se puede obtener una imagen 3D normalizada que se puede visualizar realizando cortes o promediando en distintos planos como en la imagen inferior, donde al igual que en anteriores representaciones, se ha marcado en rojo la ubicación de cada blanco (*Figura 37*).



*Figura 37: Proyecciones de una Imagen 3D  $P_{Persona}(\theta, \varphi, r)$  generada a partir de 7 blancos sintéticos*

## 4.2 MODELO COMPLEJO 3D

Al utilizar el modelo anterior en aplicaciones reales, surge la necesidad de un modelo más preciso que represente mejor la realidad.

### 4.2.1 NECESIDAD DE UN MODELO MÁS RIGUROSO

Para evaluar el funcionamiento del modelo real planteado, se realizó un experimento en el que se colgaron 7 balones en la cámara anecoica simulando el modelo de 7 blancos planteado en el apartado anterior (*Tabla 3*) aunque con las rodillas 1° más juntas entre sí. A continuación, se muestra un croquis del escenario planteado (*Figura 38*) y una imagen óptica tomada a través de la cámara en el array mientras se realizaba una captura (*Figura 39*).

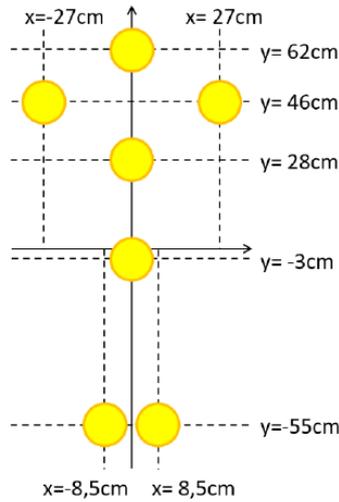


Figura 38: Esquema de los 7 balones

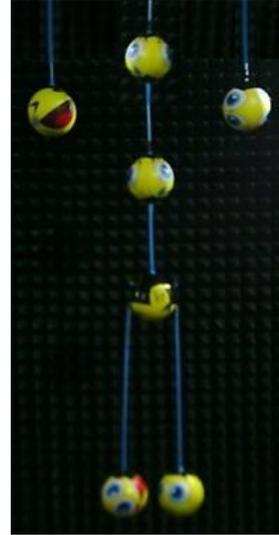
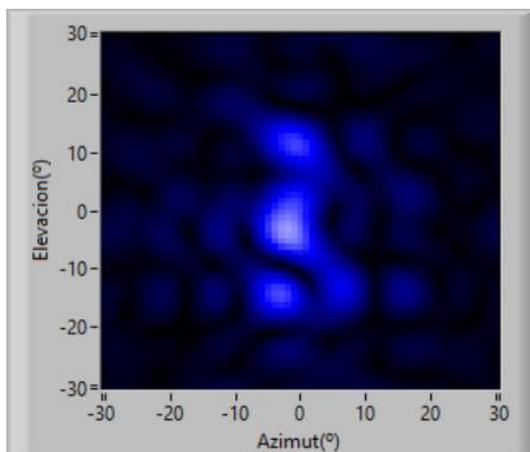
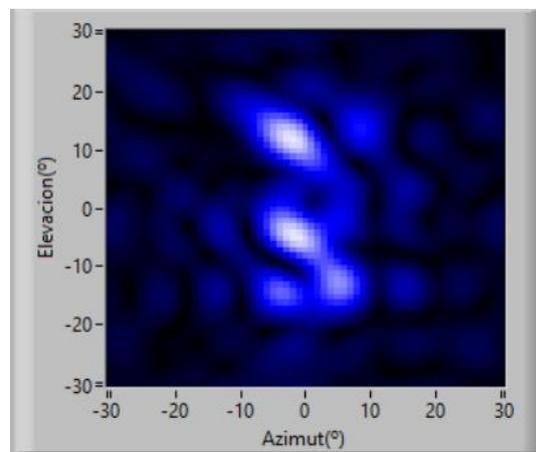


Figura 39: Imagen óptica de los 7 balones colgados

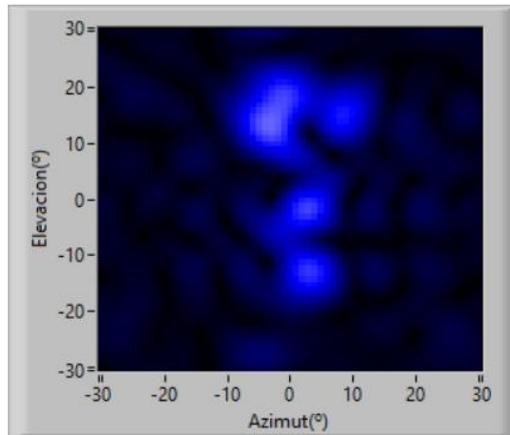
Para ver los blancos reales en las imágenes acústicas obtenidas es mejor realizar varios cortes en la imagen antes que aplicar un promedio como en otras ocasiones. En rango se puede ver como primero aparece principalmente el balón central, aunque también se intuyen los balones de rodillas y pecho (*Figura 40a*). Después se puede ir viendo como los balones de las piernas y de los hombros (más alejados en rango), empiezan a aparecer (*Figura 40b*). Finalmente, en la última imagen se ve el blanco asociado a la cabeza (*Figura 40c*).



(a) Corte en  $r = 184$



(b) Corte en  $r = 190$



(c) Corte en  $r = 200$

Figura 40: Cortes en distintos planos acimut-elevación (a,b,c) de la imagen acústica capturada de las 7 bolas

Después se compararon las capturas reales con los modelos simulados a partir de la posición real de los blancos. Fue entonces cuando se vio que en términos de error cuadrático medio (ECM), la imagen real era más similar al modelo en algunas otras disposiciones de los blancos que no correspondían a las posiciones reales. Este problema se achaca a un modelo incorrecto por lo que se empezó a pensar en estrategias para mejorarlo.

La imagen que ofrecía mejores valores de ECM ubicaba los blancos de la siguiente manera (Figura 41):

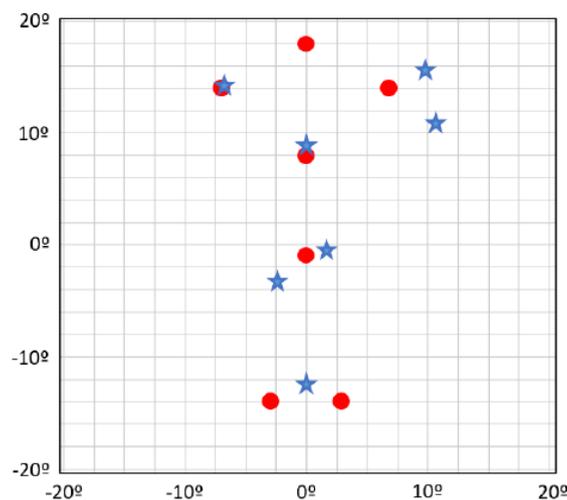


Figura 41: Posiciones: real de las bolas (círculos rojos) y ofrecidas por el modelo (estrellas azules)

Interpretar un único blanco en las piernas donde en realidad había dos es debido a la baja resolución angular del array. Sin embargo, la pelota que representa la cabeza en la parte superior tampoco es detectada por el modelo.

La causa de esto es de la sencillez del modelo. Al sumar los módulos de las respuestas espaciales (Ecuación 4.9), se está obviando el carácter complejo del diagrama de radiación y por tanto no se tienen en cuenta las fases. Esto, si el número de blancos presentes en la



$$F_i(\theta, \varphi) = \sum_{m=0}^M \sum_{n=0}^N e^{j(n-1)\gamma_\theta} \cdot e^{j(m-1)\gamma_\varphi} \quad (4.11)$$

El parámetro  $\gamma$  que aparece en las ecuaciones anteriores se puede calcular a partir de la siguiente expresión:

$$\gamma_\theta = 2\pi \cdot f \cdot \frac{d}{c} \cdot (\sin \theta_i - \sin \theta), \gamma_\varphi = 2\pi \cdot f \cdot \frac{d}{c} \cdot (\sin \varphi_i - \sin \varphi) \quad (4.12)$$

Siendo como de costumbre,  $f$ ,  $d$ ,  $c$  y  $\theta_i, \varphi_i$  la frecuencia, distancia entre sensores, velocidad del sonido y los ángulos de apuntamiento en acimut y elevación.

De esta manera es sencillo también eliminar la respuesta de los sensores erróneos o desactivados o incluso compensar las diferencias de ganancia entre algunos de ellos. Simplemente hay que utilizar una variable ( $w_{n,m}$ ) que atribuya un peso a la respuesta del sensor (siendo 0 en los casos en los que la respuesta del micrófono sea nula):

$$F_i(\theta, \varphi) = \sum_{m=0}^M \sum_{n=0}^N e^{j(n-1)\gamma_\theta} \cdot e^{j(m-1)\gamma_\varphi} \cdot w_{n,m} \quad (4.13)$$

Así pues, la imagen obtenida por el array ante un blanco  $i$  es la siguiente:

$$F_i(\theta, \varphi, r) = k_i \cdot \left( \sum_{m=0}^M \sum_{n=0}^N e^{j(n-1)\gamma_\theta} \cdot e^{j(m-1)\gamma_\varphi} \cdot w_{n,m} \right) \cdot G_{pulse}(r, r_i) \quad (4.14)$$

Al igual que antes, se calcula la respuesta del sistema ante un número  $B$  de blancos sumando las respuestas ante cada uno de ellos. Hay que tener en cuenta, no obstante, que, en esta ocasión, los blancos se suman de manera compleja para calcular el módulo una vez sumadas las respuestas de todos ellos y no antes.

$$P_{Persona}(\theta, \varphi, r) = \left| \sum_{i=1}^B F_i(\theta, \varphi, r) \right| \quad (4.15)$$

También, aunque no está recogido en las fórmulas, la imagen se normaliza por su valor máximo por comodidad.

Para finalizar, como se ha anticipado previamente, el modelo permite tener en cuenta la ganancia individual de cada sensor (Figura 43). El software de adquisición puede aproximar estadísticamente estos valores por lo que esta información se puede utilizar en el modelo.

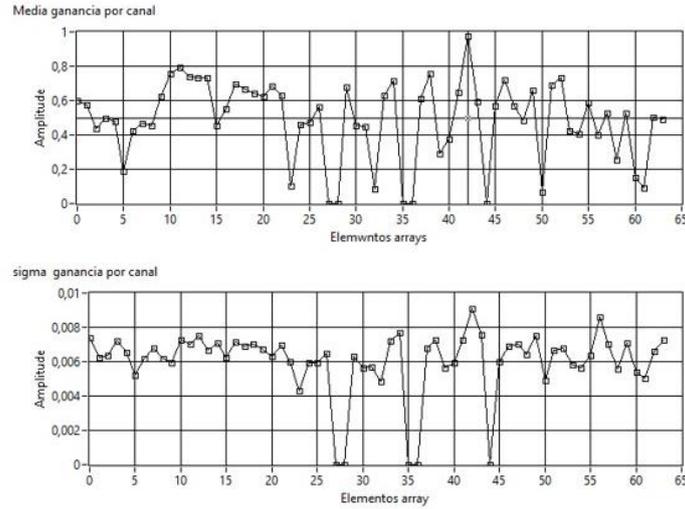


Figura 43: Medidas de media y desviación típica en la ganancia de cada uno de los micrófonos MEMS proporcionadas por el sistema de adquisición

### 4.2.3 ANÁLISIS DEL NUEVO MODELO

Se repite a continuación el experimento de los 7 balones colgados y se comparan los resultados obtenidos a través del modelo complejo (Figura 44b) con los que se obtuvieron a través del modelo real (Figura 44a).

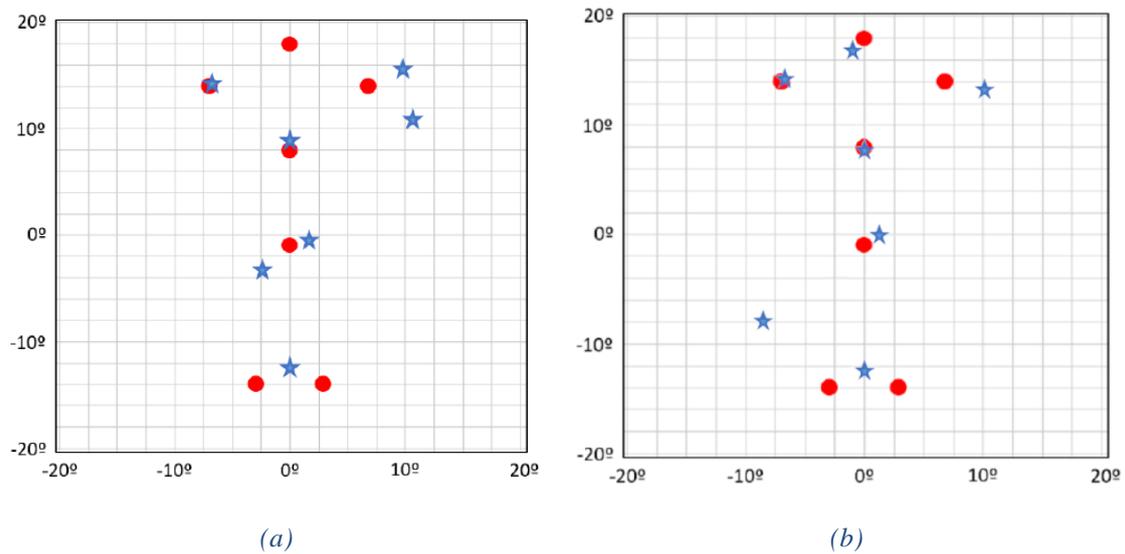


Figura 44: Comparativa de resultados obtenidos para el experimento planteado con (a) el modelo simple y (b) el modelo complejo. Posiciones: real de las bolas (círculos rojos) y ofrecidas por el modelo (estrellas azules)

De este modo, a costa de aumentar notablemente el coste computacional del modelo, se ha logrado incrementar notablemente su exactitud. Por otra parte, a pesar de haber logrado un modelo más preciso, la resolución angular del array no ha aumentado por lo que se siguen sin distinguir los balones que representan las rodillas.

Otro punto de discusión es el orden del modelo. Para el experimento realizado, es coherente asumir que la imagen recibida está compuesta por 7 blancos puesto que los

balones están suficientemente espaciados entre sí como para que así sea, pero la superficie de un individuo es mucho más compleja por lo que el orden del modelo puede no estar tan claro [8].

### 4.3 MODELO INTELIGENTE DE MÁXIMOS 3D

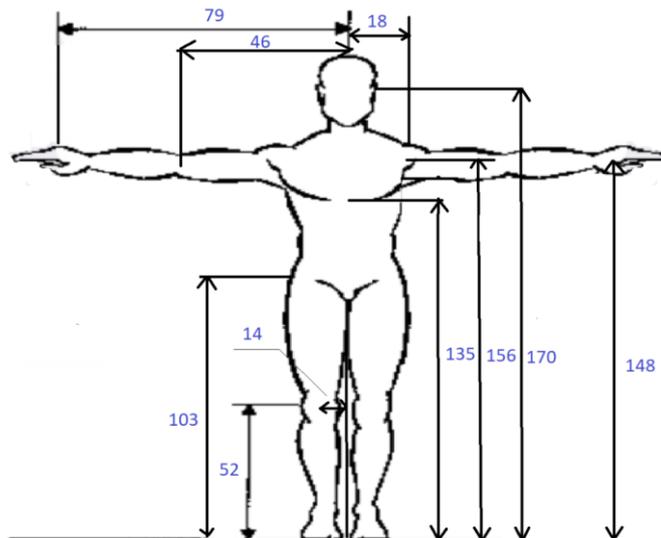
Por último, se plantea un tercer modelo a través del cual se pretende lograr una mejor representación de las características biométricas del individuo capturado. Se asume para ello que cada uno de los blancos que caracterizan a una persona, se encuentra ubicado en un máximo de la imagen.

No hay que olvidar que el objetivo principal del modelo es escoger una serie de parámetros suficientemente significativos del individuo que sean de calidad para su uso biométrico. Así, se antepone la búsqueda de un modelo que represente no tanto la imagen en sí, sino las reflexiones principales producidas por la persona.

#### 4.3.1 ESTUDIO DE LOS BLANCOS SIGNIFICATIVOS DE UN INDIVIDUO

Si el objetivo del modelo es representar las reflexiones acústicas principales de una persona, será pertinente estudiar primero cuáles son y donde se encuentran ubicadas.

Para ello, se ha estudiado un perfil en concreto sobre el cual se han tomado distintas medidas antropométricas (*Figura 45*).



*Figura 45: Medidas antropométricas tomadas sobre el individuo*

A partir de estas medidas y teniendo en cuenta las características del sistema de adquisición, se ha construido un esquema que muestra la posición de algunos puntos significativos de la figura antropométrica del individuo (cabeza, pecho, cintura, hombros, codos, manos, muslos, rodillas y tobillos) (*Figura 46*).

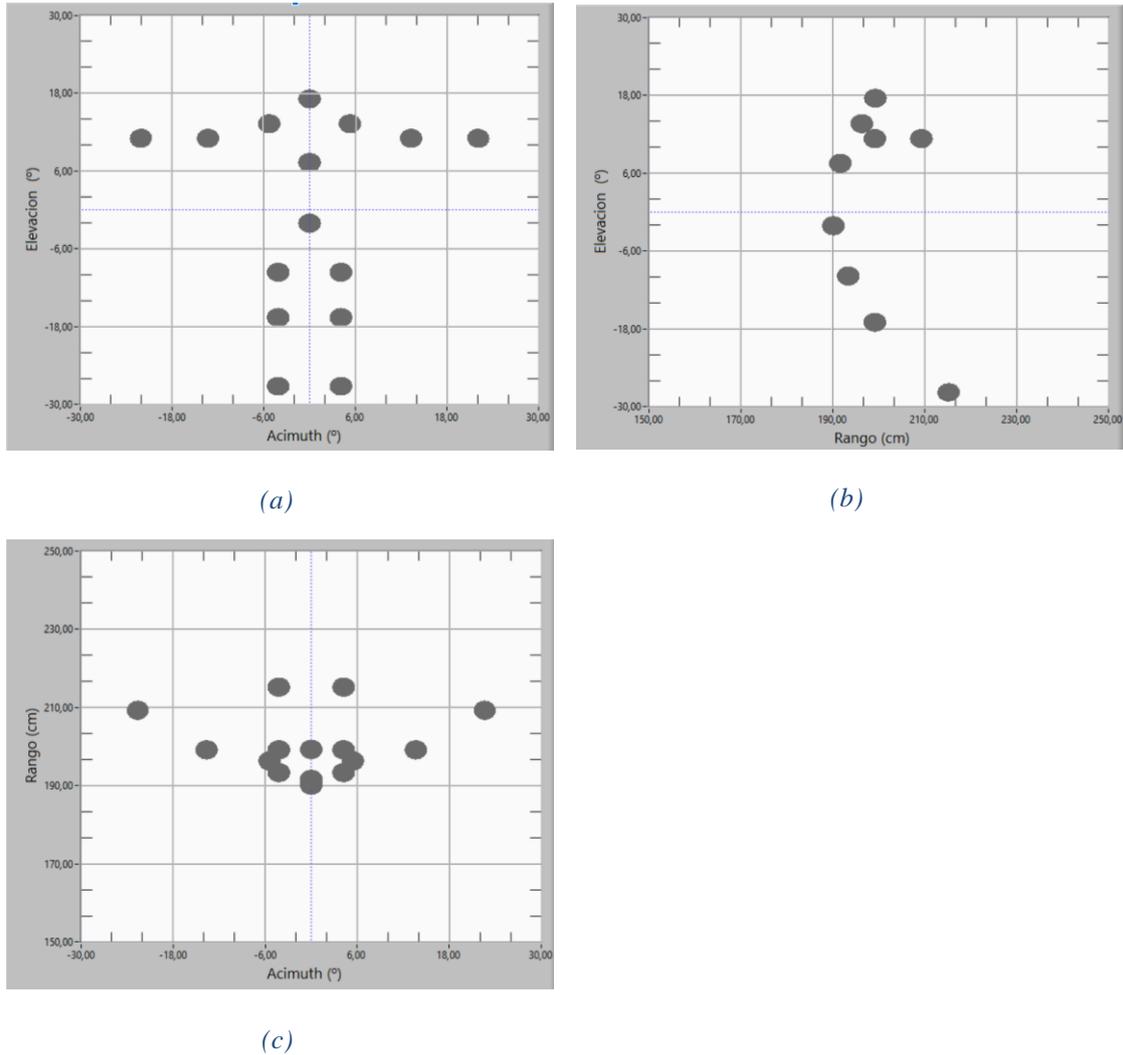


Figura 46: Esquema de los puntos significativos del individuo medido: (a) Vista frontal, (b) vista lateral y (c) vista superior

El siguiente paso ha sido calcular los máximos presentes en las capturas acústicas tomadas al individuo y compararlos con el esquema elaborado a partir de sus medidas antropométricas. En las imágenes a continuación, los círculos grises representan los puntos del esquema antropométrico, mientras que las equis verdes representan un máximo en la imagen capturada (de mayor intensidad cuando se encuentran rodeadas por una circunferencia roja).

Normalmente, los 2 máximos más fuertes corresponden a cabeza y cintura seguidos del pecho que puede o no salir como un máximo en la imagen (posiblemente debido a la complejidad reflectiva de la zona) (Figura 47).

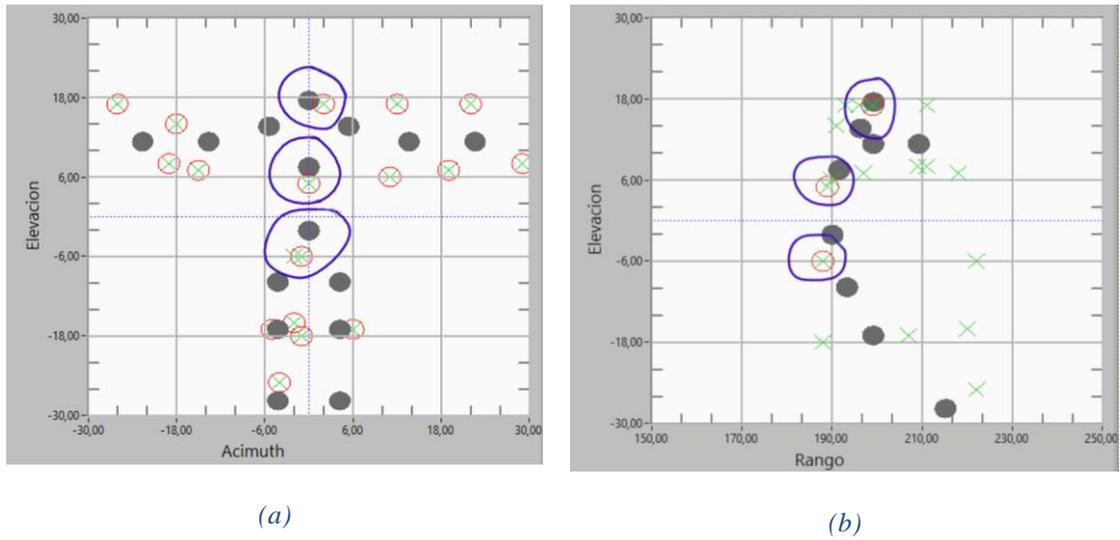


Figura 47: Máximos correspondientes a cabeza, pecho y cintura. (a) vista frontal y (b) vista lateral

Respecto a los brazos, los máximos se encuentran normalmente más dispersos, aunque se ha identificado un patrón típico de reflexión en manos y también en lo que parece ser la zona comprendida entre los codos y los hombros (Figura 48).

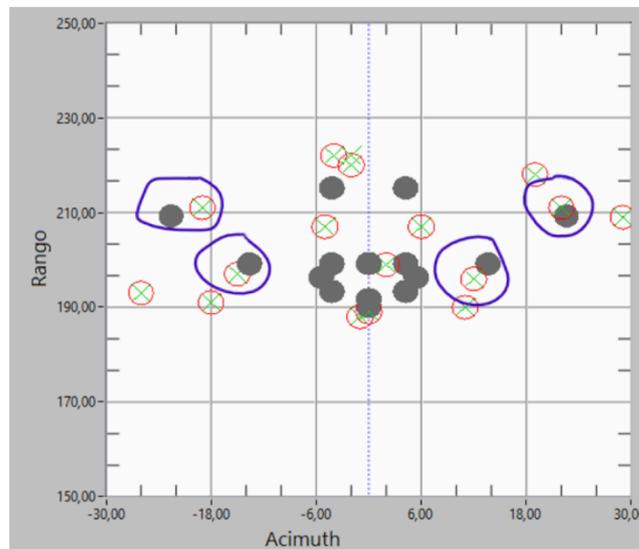


Figura 48: Máximos correspondientes a brazos y manos (Vista superior)

Sobre las extremidades inferiores, típicamente se encuentran dos máximos entre la zona de las rodillas y la de los muslos (Figura 49).

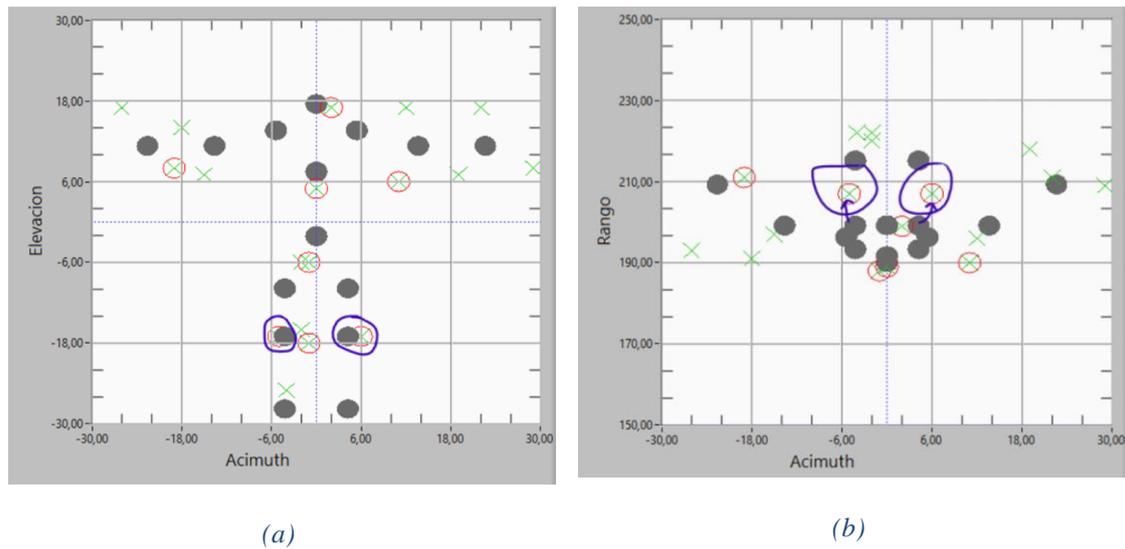


Figura 49: Máximos correspondientes a las rodillas. (a) vista frontal y (b) vista lateral

También en algunas ocasiones, aunque no tan claramente, se pueden recibir uno o dos máximos originados por la parte inferior de las piernas (zona de la tibia, tobillos o pies).

Debido a micro movimientos de la persona y cambios de postura la posición de estos blancos varía de unas capturas a otras por lo que lo importante es identificar no el punto exacto sino las zonas donde estos máximos se suelen dar.

Por último, hay que destacar que debido a la gran influencia de los lóbulos secundarios de los blancos más reflectantes de la imagen (típicamente la cintura y cabeza), escoger los  $n$  máximos más fuertes de la imagen no sería una buena opción. Es por ello por lo que para realizar el análisis anterior se han seleccionado los máximos de la imagen a lo largo de unas zonas susceptibles de poder formar parte de reflexiones acústicas de la persona.

En la imagen a continuación se rodean 6 máximos que corresponden a lóbulos secundarios de cabeza, pecho y cintura (Figura 50).

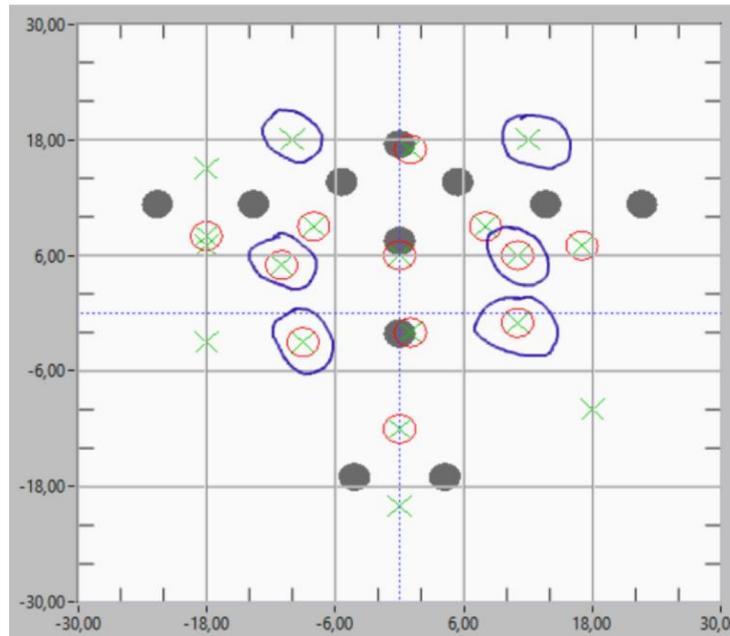


Figura 50. Máximos asociados a lóbulos secundarios. Vista frontal

La clave para distinguir un máximo perteneciente a un blanco propio de un lóbulo secundario es utilizar la información en rango. En esencia, si un máximo es un lóbulo secundario de otro máximo, ambos deben encontrarse aproximadamente en el mismo valor de rango.

#### 4.3.2 PRESENTACIÓN DEL MODELO INTELIGENTE

De este modo, se ha elaborado un modelo construido a partir de 11 máximos significativos de la imagen:

1. Cintura
2. Cabeza
3. Pecho
4. Brazo izquierdo
5. Brazo derecho
6. Mano izquierda
7. Mano derecha
8. Pierna superior izquierda
9. Pierna superior derecha
10. Pierna inferior izquierda
11. Pierna inferior derecha

Debido a la gran variabilidad presente de unas imágenes a otras, estos máximos pueden no estar presentes en alguna captura, en este caso se considera el blanco no presente con un valor nulo de amplitud indicando así que no se encuentra presente en la imagen.

#### 4.4 CONSIDERACIONES 4D

En todos los modelos propuestos se plantea trabajar únicamente a la frecuencia de  $14Khz$ . De este modo, se está dejando de capturar información a otras frecuencias que podría ser relevante.

Una posible idea sería plantear un modelo (ya sea real o complejo) que represente de manera paralela la respuesta espacial del array en distintas frecuencias. De este modo, dado que los posibles blancos de una persona están supuestamente presentes independientemente de la frecuencia, se puede buscar la posición de los blancos que mejor responda no a uno sino a todos los modelos en paralelo.

Sin embargo, la frecuencia afecta a la manera en la que el cuerpo refleja las ondas acústicas y por supuesto a la resolución del array. De esta forma, a bajas frecuencias quizás ni siquiera los blancos sean distinguibles o si lo son quizás a esas frecuencias las reflexiones se reciban más fuertes por parte de otros elementos del cuerpo.

Las imágenes multifrecuenciales se capturan enviando un pulso acústico en el que se transmiten varios tonos. Esto permite analizar la respuesta del individuo capturado a distintas frecuencias. Este pulso multitono se genera a través de una técnica LUT (*Look Up Table*). Es decir, el software construye el pulso deseado a partir de las frecuencias escogidas y lo almacena en una tabla en memoria (LUT). De este modo la myRIO lee los valores de la tabla para transmitir así el pulso personalizado.

# 5

## Capítulo 5. EXTRACCIÓN DE PARÁMETROS

---

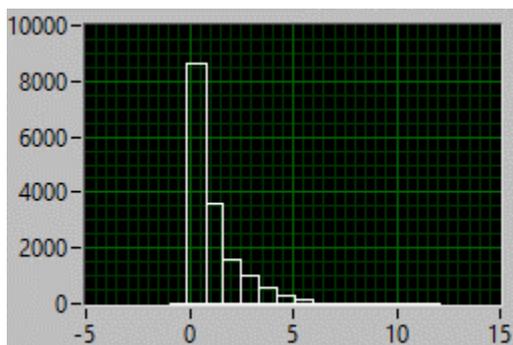
El siguiente paso después de plantear un modelo es diseñar una estrategia para obtener los parámetros que permiten construirlo. Estos parámetros permiten construir las características que se utilizan posteriormente en la etapa de *machine learning* por lo que es importante que el error en la estimación de los parámetros sea reducido.

En el caso de los dos primeros modelos, la estrategia para obtener los parámetros que construyen el modelo consiste en generar a partir del modelo matemático una imagen sintética que simula la imagen real capturada. Así, se obtienen los parámetros estimados como aquellos que minimizan las diferencias entre la imagen real y una nueva imagen sintética construida a partir de dichos parámetros estimados.

Para el modelo inteligente de máximos 3D la estrategia (que se comentará más adelante en “5.3 Algoritmo de Búsqueda Inteligente”) será ligeramente diferente dado que la filosofía principal de este modelo no es construir una imagen lo más parecida a la real, sino obtener unos parámetros de calidad.

La calidad de la estrategia de extracción de parámetros empleada se puede evaluar o bien a través de imágenes reales o bien por medio de simulaciones. Una imagen simulada se construye a partir del modelo planteado, más una componente de ruido aleatorio que emula todo aquello que el modelo no es capaz de predecir.

Para caracterizar el ruido acústico presente en las imágenes reales, se han tomado capturas sin introducir ningún tipo de elemento de reflexión en la cámara anecoica y se ha obtenido el histograma del ruido (*Figura 51*). Esto permite realizar una primera aproximación grosso modo.



*Figura 51: Histograma del ruido presente en una imagen acústica*

A pesar del carácter exponencial o gamma que se puede intuir en la aproximación realizada sobre el ruido, en las simulaciones se ha utilizado generalmente ruido gaussiano,

de media nula, y desviación estándar  $\sigma_N$ . Se ha escogido el ruido gaussiano para las simulaciones ya que su efecto es fácil de modelar y cuantificar. Además, las principales fuentes de ruido en las imágenes acústicas serán otras reflexiones indeseadas por lo que representar de manera exacta el ruido de fondo no es necesario.

## 5.1 PLANTEAMIENTO INICIAL DE ESTIMACIÓN. LSE

Para plantear la estimación de los parámetros, se considera la imagen real o simulación  $P_{persona}(\theta, \varphi, r)$  como un registro 3D de observaciones muestreadas  $x[\theta, \varphi, r]$  con  $\theta \in [-30^\circ, 30^\circ]$ ,  $\varphi \in [-30^\circ, 30^\circ]$  y  $r \in [150cm, 250cm]$ .

Por otra parte,  $s[\theta, \varphi, r; \boldsymbol{\rho}]$  representa el modelo paramétrico escogido de la imagen. En el caso de la simulación, la componente de ruido aleatorio aditivo de la imagen corresponde con lo que se denominará a partir de ahora como  $w[\theta, \varphi, r]$ , mientras que para las imágenes reales  $w[\theta, \varphi, r]$  representa todo aquello que el modelo no puede representar como el propio ruido de medida.

De esta manera, las observaciones pueden escribirse de la siguiente forma: (Ecuación 5.1)

$$x[\theta, \varphi, r] = s[\theta, \varphi, r; \boldsymbol{\rho}] + w[\theta, \varphi, r; \boldsymbol{\rho}] \quad (5.1)$$

Los parámetros que componen el modelo paramétrico de la imagen se recogen en el vector  $\boldsymbol{\rho}$ , el cual corresponde al conjunto de  $p$  parámetros que se quieren estimar. A modo de ejemplo, para una imagen formada por 3 blancos,  $p = 12$  parámetros a estimar y  $\boldsymbol{\rho} = [k_1, \theta_1, \varphi_1, r_1, k_2, \theta_2, \varphi_2, r_2, k_3, \theta_3, \varphi_3, r_3]$ .

Se plantea el problema entonces bajo el enfoque de estimación clásica, es decir, asumiendo que el valor de  $\boldsymbol{\rho}$  corresponde a una constante determinística desconocida y sin conocer información a priori. Con esta idea y tras el estudio de varios estimadores, se ha considerado plantear el problema como uno de estimación por mínimos cuadrados (LSE, *least Square estimator*).

El concepto de partida es considerar el modelo de señal diseñado  $s[\theta, \varphi, r; \boldsymbol{\rho}]$  como apropiado. De esta manera, la componente de ruido  $w[\theta, \varphi, r]$  se tiene en cuenta únicamente como una imperfección del modelo y sin asumir nada sobre ella.

Esto resulta ser un aspecto positivo pues a pesar de que en el modelo simulado se puede caracterizar perfectamente  $w[\theta, \varphi, r]$ , para las imágenes reales no es tan sencillo. Además, el método de mínimos cuadrados es completamente determinístico, por lo que no será necesario manejar funciones de densidad de probabilidad en ningún momento.

De esta manera, se define una función de costes  $J(\boldsymbol{\rho})$  que resulta de la suma de las diferencias al cuadrado entre la imagen simulada  $x[\theta, \varphi, r]$  y el modelo paramétrico  $s[\theta, \varphi, r; \boldsymbol{\rho}]$ .

$$J(\boldsymbol{\rho}) = \sum_{\theta=-30}^{30} (x[\theta, \varphi, r] - s[\theta, \varphi, r; \boldsymbol{\rho}])^2 \quad (5.2)$$

Nótese que los errores se encuentran elevados al cuadrado con el objetivo de evitar cancelaciones de signo entre los que presentan valores negativos y los positivos.

Esta función de costes representa la suma de todos los errores cometidos. Por ello,  $J(\boldsymbol{\rho})$  se puede interpretar como una medida de calidad de los parámetros calculados ya que, cuánto más parecidos sean a los que originaron la observación, más similares serán  $x[\theta, \varphi, r]$  y  $s[\theta, \varphi, r; \boldsymbol{\rho}]$  y menor será el valor de  $J(\boldsymbol{\rho})$ . Por consiguiente, si el modelo paramétrico es bueno, el estimador también lo será.

Sobre esta base conceptual se define el estimador de mínimos cuadrados  $\hat{\boldsymbol{\rho}}_{LSE}$ .

$$\hat{\boldsymbol{\rho}}_{LSE} = \arg \min_{\boldsymbol{\rho}} J(\boldsymbol{\rho}) = \arg \min_{\boldsymbol{\rho}} \sum_{r=150}^{250} \sum_{\varphi=-30}^{30} \sum_{\theta=-30}^{30} (x[\theta, \varphi, r] - s[\theta, \varphi, r; \boldsymbol{\rho}])^2 \quad (5.3)$$

La solución es sencilla y conocida para un caso lineal, sin embargo, tanto en el modelo real como en el modelo complejo  $s[\theta, \varphi, r; \boldsymbol{\rho}]$  se encuentran muy lejos de depender linealmente de los parámetros a estimar por lo que hallar la solución es complejo. De hecho, debe hablarse del estimador como un NLSE (*Non-Linear Least-Squares Estimator*).

En base al planteamiento sugerido, para encontrar el estimador  $\hat{\boldsymbol{\rho}}_{LSE}$ , el problema de estimación se transforma en un problema de optimización no lineal. Desde este punto de vista, la función objetivo que se pretende minimizar es  $J(\boldsymbol{\rho})$  y las variables de optimización corresponden a los parámetros  $\boldsymbol{\rho}$ .

A modo de ejemplo, se ha planteado un escenario simplificado. Se define un array lineal (donde la única variable es el acimut) y se calcula la respuesta espacial de un único blanco de amplitud conocida, pero ángulo desconocido. Según la estrategia planteada, aquel valor de  $\theta$  para el cuál la función de costes definida en la (Ecuación 5.2) tome su valor mínimo será el ángulo donde se encuentre el blanco.

Se ha programado a través de MATLAB tanto el modelo (Figura 52) como el cálculo exhaustivo de la función de costes para todos los valores de acimut posibles. Se constata que dado que el blanco está ubicado en  $\theta = 0$ , el valor mínimo de la función de costes se alcanza en este valor (Figura 53).

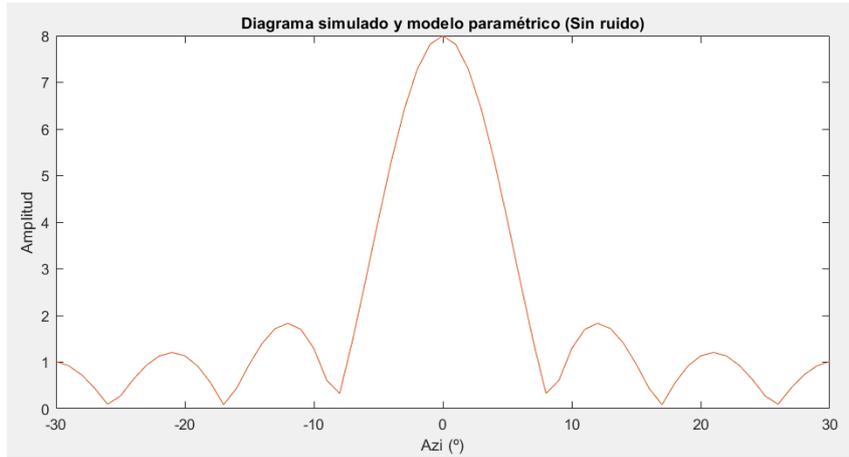


Figura 52: Señal recibida y estimación del modelo para un único blanco. Caso sin ruido

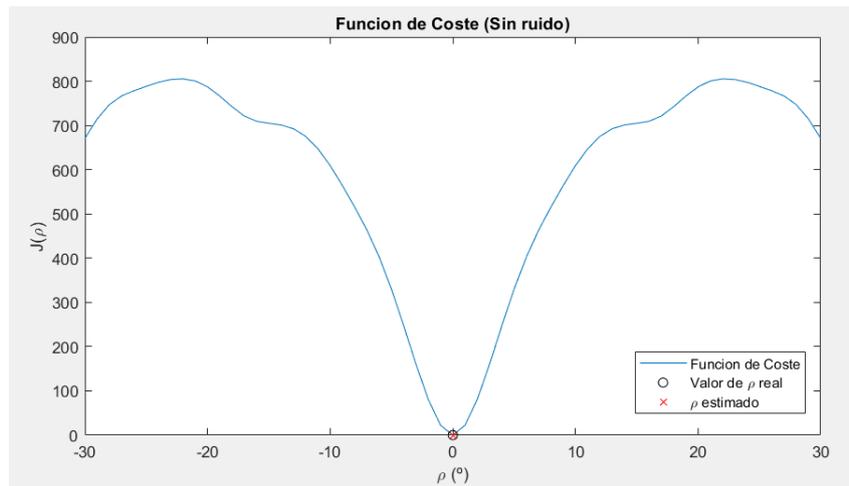


Figura 53: Función de costes para un único blanco. Caso sin ruido

Si se agrega a la señal simulada un ruido exponencial con múltiples varianzas, se puede comprobar la robustez del modelo. De hecho, para el caso con  $\sigma = 6$ , el diagrama de radiación recibido (señal azul en (Figura 54)) al contar con una componente de ruido elevada, se aleja notablemente del diagrama de radiación estimado, pero a pesar de ello, la función de costes presenta un mínimo absoluto lo suficientemente claro (Figura 55).

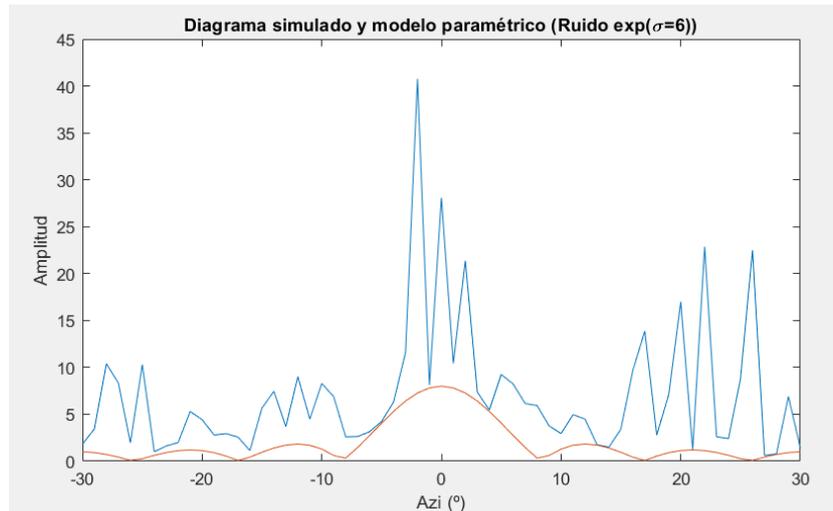


Figura 54: Señal recibida y estimación del modelo para un único blanco contaminado con ruido exponencial de  $\sigma = 6$

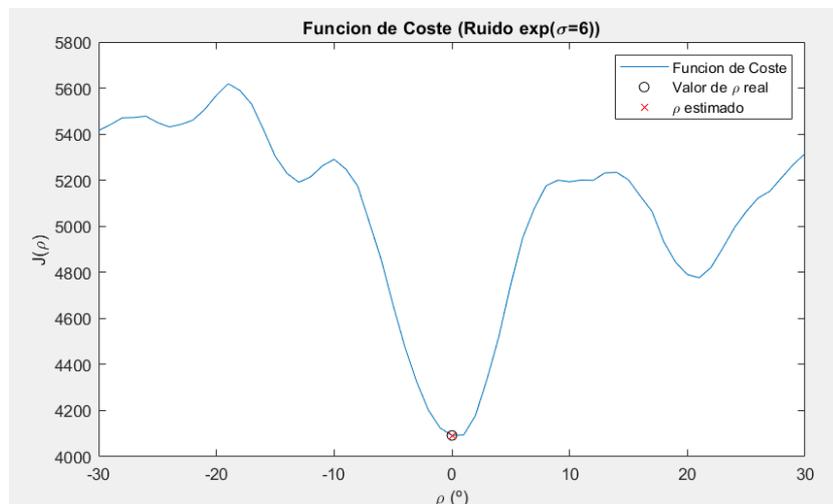


Figura 55: Función de costes para un único blanco contaminado con ruido exponencial de  $\sigma = 6$

Por último, pueden obtenerse dos conclusiones a partir de estos resultados:

1. Los mínimos relativos de la función de costes pueden ser un problema para algunos algoritmos de estimación al tomar este mínimo local como mínimo absoluto.
2. Cuando la dimensionalidad del problema es elevada, no se puede analizar de forma visual el comportamiento de los mínimos locales.

## 5.2 OPTIMIZACIÓN. ALGORITMOS DE BÚSQUEDA

La búsqueda de la solución óptima generalmente consiste en un proceso iterativo. Se generan una sucesión de puntos sobre un espacio de búsqueda en los que cada uno se obtiene a partir de una transformación sobre el anterior. Sin embargo, existen numerosas formas de llevar a cabo esta búsqueda que presentan distinta eficacia acorde a la situación

planteada. Para entender la problemática y aplicado al caso de estudio particular, se plantean a continuación 2 posibles soluciones heurísticas.

### 5.2.1 BÚSQUEDA EXHAUSTIVA

#### **Planteamiento**

Para estar seguros de encontrar el valor óptimo en el espacio de búsqueda se podría evaluar una a una todas las posibles soluciones o puntos del espacio. Este procedimiento recibe el nombre de “búsqueda exhaustiva” y proporciona la certeza de que la solución encontrada es la mejor para el espacio de búsqueda [22].

No obstante, que la solución encontrada a través de una búsqueda sea óptima para ese espacio de búsqueda no garantiza que sea el óptimo global, pues podría tratarse de un mínimo local que sea absoluto para ese espacio (de hecho, esto se conoce como búsqueda local).

Así pues, la elección del espacio de búsqueda es un factor importante para tener en cuenta, visto que, si la solución global no está contenida en el espacio de búsqueda, ningún algoritmo será capaz de encontrarlo.

En consecuencia, si se realiza la búsqueda por todo el espacio global, o, dicho de otra manera, el espacio de búsqueda corresponde al global, se podría afirmar que necesariamente uno de esos puntos es el óptimo global.

Sin embargo, el coste computacional de probar todas y cada una de las soluciones hace inviable el método para problemas reales donde el espacio global es sumamente grande.

#### **Procedimiento**

Se considera como ejemplo únicamente un escenario de 3 blancos capturado mediante un array lineal siendo el acimut la variable. De este modo  $\boldsymbol{\rho} = [k_1, \theta_1, k_2, \theta_2, k_3, \theta_3]$  y el caso de estudio sería el siguiente:

Se considera que  $\theta_i$  puede tomar 61 valores, puesto que  $\theta_i \in [-30^\circ, 30^\circ]$  y se barren valores discretos con resolución de  $1^\circ$ . También, suponiendo la imagen normalizada respecto al blanco máximo y una resolución en amplitud de 0,1,  $k_i$  puede tomar 10 valores posibles (desde 0,1 hasta 1). Es decir, el espacio global cuenta con  $10^3 \cdot 61^3 = 226.981.000$  puntos que el algoritmo recorrerá de uno a uno.

El procedimiento, como en todos los algoritmos de búsqueda exhaustiva, es simple. Se evalúa el valor de la función objetivo  $J(\boldsymbol{\rho})$  para todas y cada una de las 226.981.000 posibles soluciones. Se va almacenando el valor mínimo de la función objetivo respecto a todos los valores obtenidos en el camino recorrido y, siempre que ese valor sea superado por un punto nuevo, se actualiza el valor mínimo y el estimador  $\hat{\boldsymbol{\rho}}$  que corresponde a ese punto.

De esta forma, al finalizar, el programa tendrá almacenado  $\hat{\boldsymbol{\rho}}$  como aquel conjunto de parámetros de todos los posibles que minimiza de manera global  $J(\boldsymbol{\rho})$ .

### Ejemplo práctico. Análisis del algoritmo

Se ha generado una imagen contaminada con ruido gaussiano de varianza  $\sigma = 0,5$  simulada a través del modelo simple.

Después se ha programado el algoritmo para el ejemplo de 3 blancos planteado en el apartado 4.1.1 “Síntesis 1D“, cuyos parámetros se recogen en la (Tabla 1). Tras una media de  $2h$  y  $2min$ , el algoritmo finaliza la búsqueda y proporciona la mejor solución que coincide con los parámetros originales (Figura 56).

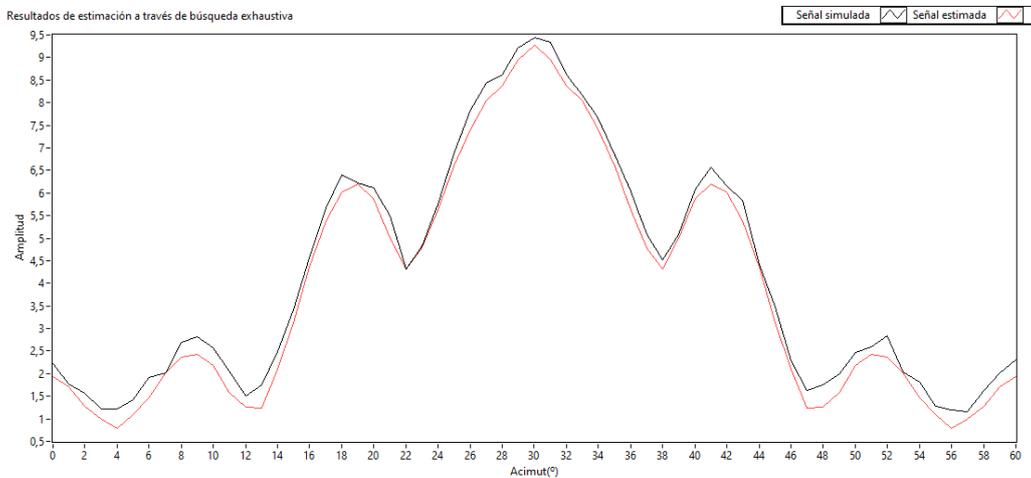


Figura 56: Resultados de la búsqueda exhaustiva aplicada sobre el modelo de 3 blancos

Planteando un caso real de imágenes acústicas 3D en el que se consideren 11 blancos,  $\rho = [k_1, \theta_1, \varphi_1, r_1, k_2, \theta_2, \varphi_2, r_2, \dots, k_{11}, \theta_{11}, \varphi_{11}, r_{11}]$ , por lo que  $p = 44$  parámetros a estimar. Esto, contando con que tanto  $\theta_i$  como  $\varphi_i$  pueden tomar 61 valores,  $r_i$  puede tomar 101 valores (entre 150cm y 250cm) y  $k_i$  10 (según lo considerado en el anterior apartado), hace que el espacio global de soluciones este formado por  $10^{11} \cdot 61^{11} \cdot 61^{11} \cdot 101^{11} \approx 2,11 \cdot 10^{72}$  puntos posibles. El número de soluciones posibles es muy elevado, siendo este método inviable.

### 5.2.2 BÚSQUEDA ALEATORIA

#### Planteamiento

Como se ha comprobado, el tiempo que lleva realizar una búsqueda exhaustiva por todas y cada una de las soluciones posibles es muy elevado o incluso inviable dependiendo de la magnitud del problema. Una aproximación al problema es que la búsqueda se realizase sólo por algunos puntos del espacio total de soluciones.

La búsqueda aleatoria se basa en recorrer de manera no determinista el espacio de soluciones obteniendo la mejor solución posible. Esto implica valorar como suficiente una solución buena que no necesariamente será la mejor, aunque por otra parte el tiempo de búsqueda se reducirá considerablemente. De hecho, a medida que aumente el tiempo de búsqueda del algoritmo, aumentarán las posibilidades de encontrar una solución mejor.

**Procedimiento**

El espacio de búsqueda es el mismo que para el caso de búsqueda exhaustiva, sin embargo, ahora se aleatorizan los valores de los parámetros a estimar en cada evaluación de la función objetivo. Al igual que antes se almacena el valor mínimo de la función objetivo respecto a los valores obtenidos en el camino recorrido. Cuando este valor sea inferior para un punto nuevo, se actualiza el valor mínimo y el estimador  $\hat{\rho}$  que corresponde a ese punto.

La búsqueda puede finalizar tras haber transcurrido una cantidad de tiempo determinada o tras haber conseguido alcanzar una solución cuyo error cometido se encuentre por debajo de una cota preestablecida. También se puede dar por finalizado el algoritmo tras haber pasado por cierto número de iteraciones posibles, aunque el efecto es similar al de tener en cuenta el tiempo transcurrido.

**Ejemplo práctico. Análisis del algoritmo**

Se ha generado de nuevo un escenario de 3 blancos idéntico al que se plantea en el apartado de ejemplo práctico en “Búsqueda exhaustiva”. Es decir:  $\theta_1 = 0^\circ, \theta_2 = 10^\circ, \theta_3 = -10^\circ, k_1 = 1, k_2 = 0,5$  y  $k_3 = 0,5$ . Fijando un error cuadrático máximo permitido de 0,3 para que el algoritmo finalice la búsqueda y realizando 10 iteraciones con semillas independientes se obtuvieron los resultados recogidos en la tabla a continuación (Tabla 4):

Número de Ejecución	ECM obtenido	Tiempo de ejecución	$\widehat{\theta}_1$	$\widehat{k}_1$	$\widehat{\theta}_2$	$\widehat{k}_2$	$\widehat{\theta}_3$	$\widehat{k}_3$
1	0,24931	69,888s	0°	0,95	10°	0,46	-10°	0,69
2	0,26341	39,764s	0°	0,92	11°	0,63	-10°	0,59
3	0,2739	5,55s	0°	0,98	10°	0,56	-11°	0,43
4	0,21102	19,521s	0°	0,91	10°	0,46	-10°	0,6
5	0,23531	5,609s	0°	0,93	10°	0,47	-9°	0,61
6	0,15646	1,215s	0°	0,96	10°	0,52	-9°	0,52
7	0,25135	4,572s	0°	0,95	11°	0,46	-10°	0,63
8	0,068657	17,264s	0°	0,97	10°	0,57	-10°	0,57
9	0,21961	15,199s	0°	1	11°	0,61	-10°	0,58
10	0,27578	9,795s	0°	0,87	10°	0,65	-10°	0,5

Tabla 4: Resultados obtenidos en 10 ejecuciones del algoritmo de búsqueda aleatoria para un error cuadrático medio máximo de 0,3

De este modo, si el problema no requiere de una gran precisión, esta es una buena solución. De hecho, la estimación obtenida en la ejecución 8 (Tabla 4) proporciona unos muy buenos resultados (Figura 57).

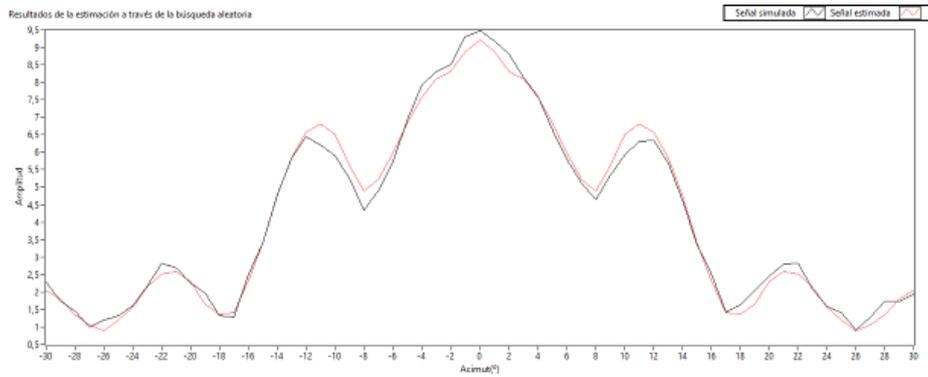


Figura 57: Mejor solución obtenida para 10 iteraciones realizadas a través de la búsqueda aleatoria fijando el error cuadrático medio máximo en 0,3

La conclusión es clara, si se requiere de una gran precisión y el tamaño del espacio de búsqueda lo permite, la solución a escoger es la búsqueda exhaustiva.

Por el contrario, si no se requiere de obtener la mejor solución y vale con una suficientemente buena, la búsqueda aleatoria resuelve en segundos un problema que la búsqueda exhaustiva resolvía en aproximadamente 2 horas. Además, este algoritmo al igual que el anterior es realmente simple y permite mayor flexibilidad que la búsqueda exhaustiva.

Aplicado a un caso de imágenes reales, el elevado número de soluciones del espacio ( $2,1124774532 \cdot 10^{72}$  para el modelo de 11 blancos) sigue dificultando cualquier tipo de búsqueda, aunque sea aleatoria. En este caso no hará falta recorrer todos los posibles puntos del espacio de búsqueda, pero la probabilidad de encontrar una solución aceptable entre tantas posibles soluciones se reduce drásticamente.

### 5.3 SQP. OPTIMIZACIÓN NO LINEAL

Dejando de lado los métodos heurísticos, uno de los métodos matemáticos más exitosos para resolver problemas de optimización no lineales sujetos a restricciones es SQP (*Sequential Quadratic Programming*). La idea básica detrás de SQP es modelar el problema a partir de una solución aproximada  $x^k$  como un subproblema de programación cuadrática y usar esta aproximación para construir una mejor aproximación  $x^{k+1}$ . Este proceso se repite para crear una secuencia de aproximaciones que se espera converjan en la solución  $x^*$ , por lo que SQP es un método iterativo [38].

Los métodos SQP comparten las características de los métodos tipo Newton, es decir, una convergencia rápida cuando los puntos iterados están cerca de la solución a costa de un posible comportamiento erróneo cuando los puntos iterados se encuentran lejos de la solución. Esto quiere decir que el método es muy sensible al punto de inicio para obtener la solución óptima por lo que la elección de los parámetros iniciales es un elemento crucial [38].

En LabVIEW, existe un módulo denominado “Constrained Nonlinear Optimization VI” (Figura 58) que permite resolver un problema de optimización no lineal a través de un método SQP [39].

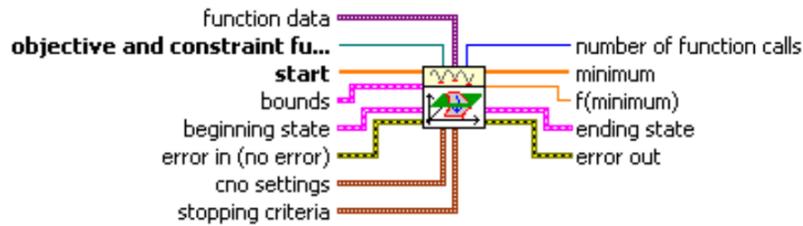


Figura 58: Constrained Nonlinear Optimization VI

De este modo, se ha elaborado una rutina en LabVIEW que, optimizando la función de costes definida a través del estimador NLSE (Ecuación 5.2) y partiendo de unos parámetros iniciales de calidad, encuentra los parámetros óptimos de la imagen.

Dada la complejidad de la función de costes y la innegable importancia del método con los parámetros iniciales, es necesaria una discusión acerca de la elección de estos.

La primera idea que se planteó es utilizar los máximos presentes en la imagen, sin embargo, como se ha comentado previamente, debido principalmente al efecto de los lóbulos secundarios, no todos los máximos correspondían a un blanco en la imagen por lo que fue necesario diseñar estrategias que escojan sólo los máximos susceptibles de corresponder a un blanco en la imagen. A través de este concepto, se ha definido un nuevo modelo que se ha denominado, “Modelo inteligente de máximos 3D” y que se detalla en la sección 5.4.

Simulando imágenes utilizando el modelo real y según el esquema de 7 bolas expuesto en anteriores apartados (Tabla 3), el algoritmo funcionó de una manera muy adecuada logrando en 1000 simulaciones un valor medio del error cuadrático medio de  $4,36 \cdot 10^{-5}$ . Estos fueron los 1000 valores estimados para cada uno de los 7 blancos (Figura 59).

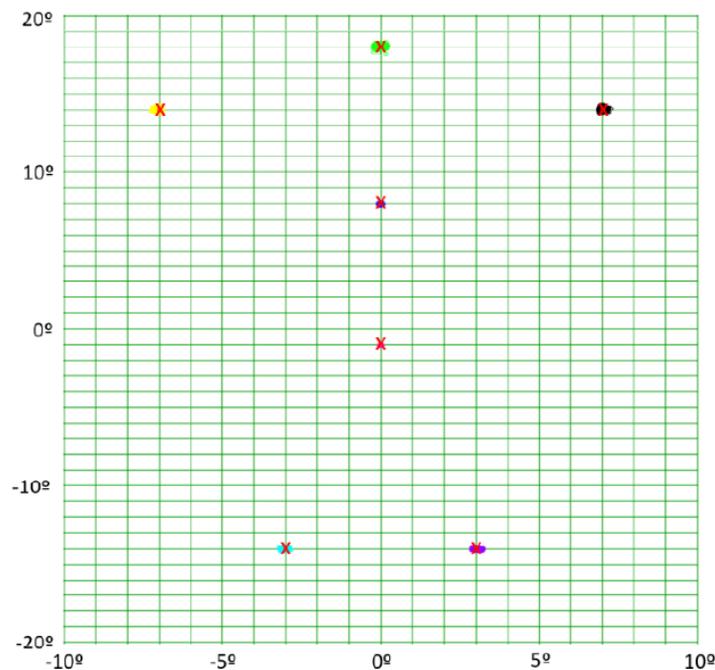


Figura 59: Representación acimut-elevación de las posiciones de los 7 blancos simulados (cruces rojas) y estimadas (puntos de colores)

En la (Figura 60) se exponen los estadísticos obtenidos a través de las 1000 simulaciones correspondientes.

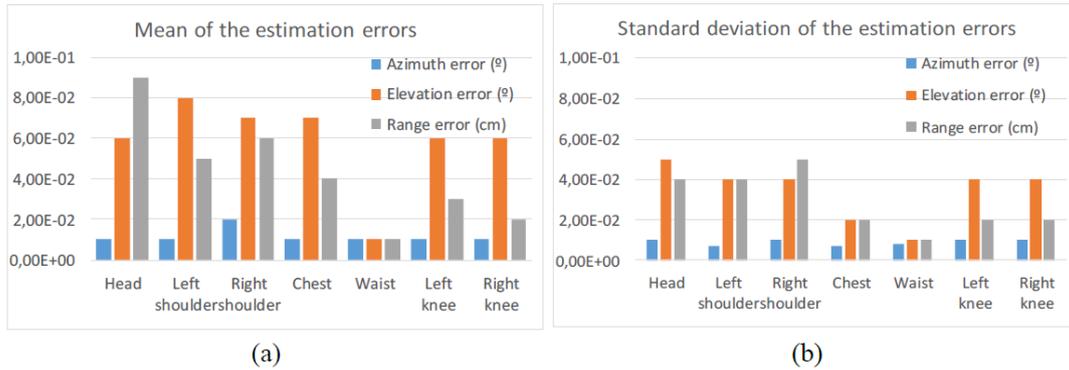


Figura 60: Parámetros estadísticos del error obtenido en la estimación de cada blanco para el escenario simulado. (a) Media, (b) Desviación estándar

Estos resultados muestran que el comportamiento del estimador no es igual para todas las partes del cuerpo simuladas. Los mayores errores se encuentran en la estimación de la cabeza y hombros y la mejor estimación se obtiene claramente para la cintura. También se puede inferir que el algoritmo muestra un mejor comportamiento a la hora de estimar la posición acimutal de los blancos. Esto es principalmente por que la gran mayoría de blancos se agrupan en el mismo valor de acimut ( $0^\circ$ ) por lo que también lo harán sus lóbulos secundarios dificultando la estimación en elevación.

En las simulaciones, se puede comprobar que el valor de los parámetros iniciales (que fueron elegidos a partir de los máximos de la imagen) se aproxima muy bien a los parámetros reales, sin embargo, cuando se utiliza en una imagen real, dada la complejidad de la misma, la posición de los máximos aporta menos información de los blancos presentes.

A pesar de que el funcionamiento del algoritmo es bueno (como se ha comprobado en simulaciones), al utilizar imágenes reales el resultado no fue tan bueno como se esperaba en un principio. Esto se debe a que las hipótesis realizadas en la elaboración del modelo no son correctas o bien los parámetros iniciales no se han elegido adecuadamente.

## 5.4 ALGORITMO DE BÚSQUEDA INTELIGENTE

En caso de utilizar el modelo inteligente de máximos 3D, es preciso identificar en las imágenes reales recibidas, cada uno de los máximos y asociarles a cada uno de los elementos de reflexión del cuerpo (cabeza, pecho, cintura, etc.). Para ello, el algoritmo se divide en dos bloques: la búsqueda de máximos en la imagen y la asociación de estos máximos a las partes del cuerpo consideradas.

Este es un algoritmo completamente heurístico y nace a partir de la experiencia sobre las observaciones realizadas durante el desarrollo de este trabajo fin de grado.

### 5.4.1 BÚSQUEDA

El algoritmo de búsqueda de máximos utiliza una estrategia basada en un núcleo (*kernel*) 3D que se evalúa sobre cada punto de la imagen, si la amplitud del punto central del

núcleo es mayor que la de todos sus elementos de alrededor, se clasifica el punto como máximo. El algoritmo es más permisivo si el tamaño del núcleo es menor por lo que se detectarían un mayor número de máximos. El ejemplo a continuación muestra un núcleo 3D que evalúa 9 valores (Figura 61). De este modo, sólo si el valor central es superior a los 8 de alrededor se considera como máximo.

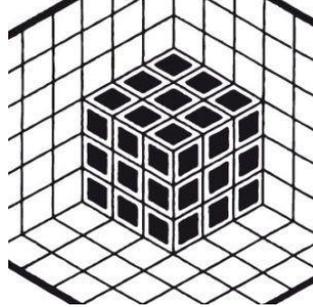


Figura 61: Ejemplo de kernel 3D (3x3x3)

De este modo, a la salida se tiene una lista de todos los máximos presentes en la imagen ordenados según su amplitud.

#### 5.4.2 ASOCIACIÓN. CONTENEDORES ADAPTATIVOS

Una vez localizados todos los máximos de la imagen el problema es identificar cuáles de ellos se asocian a las reflexiones del individuo consideradas. Para ello, se han definido un conjunto de contenedores tridimensionales de manera antropométricamente coherente para cualquier persona (tabla 5).

Contenedores:	$\theta_{min}$	$\theta_{max}$	$\varphi_{min}$	$\varphi_{max}$	$r_{min}$	$r_{max}$
Cadera ( $C_1$ )	-4	4	-6	6	180	215
Cabeza ( $C_2$ )	-4	4	$\varphi_{C_1} + 5$	27	$r_{C_1} + 1$	$r_{C_1} + 20$
Pecho	-3	3	$\varphi_{C_1} + 5$	$\varphi_{C_2} - 1$	$r_{C_1}$	$r_{C_2} - 1$
Brazo izquierdo	5	18	$\varphi_{C_1} + 5$	$\varphi_{C_2} - 4$	$r_{C_1} + 2$	$r_{C_1} + 20$
Brazo derecho	-18	-5	$\varphi_{C_1} + 5$	$\varphi_{C_2} - 4$	$r_{C_1} + 2$	$r_{C_1} + 20$
Mano izquierda	18	27	$\varphi_{C_1} + 5$	$\varphi_{C_2}$	$r_{C_1} + 10$	$r_{C_1} + 30$
Mano derecha	-27	-18	$\varphi_{C_1} + 5$	$\varphi_{C_2}$	$r_{C_1} + 10$	$r_{C_1} + 30$
Parte superior de la pierna izquierda	3	7	$\varphi_{C_1} - 15$	$\varphi_{C_1} - 4$	$r_{C_1} - 4$	$r_{C_1} + 10$
Parte superior de la pierna derecha	-7	-3	$\varphi_{C_1} - 15$	$\varphi_{C_1} - 4$	$r_{C_1} - 4$	$r_{C_1} + 10$
Parte inferior de la pierna izquierda	3	7	-29	$\varphi_{C_1} - 20$	$r_{C_1} + 10$	215
Parte inferior de la pierna izquierda	-7	-3	-29	$\varphi_{C_1} - 20$	$r_{C_1} + 10$	215

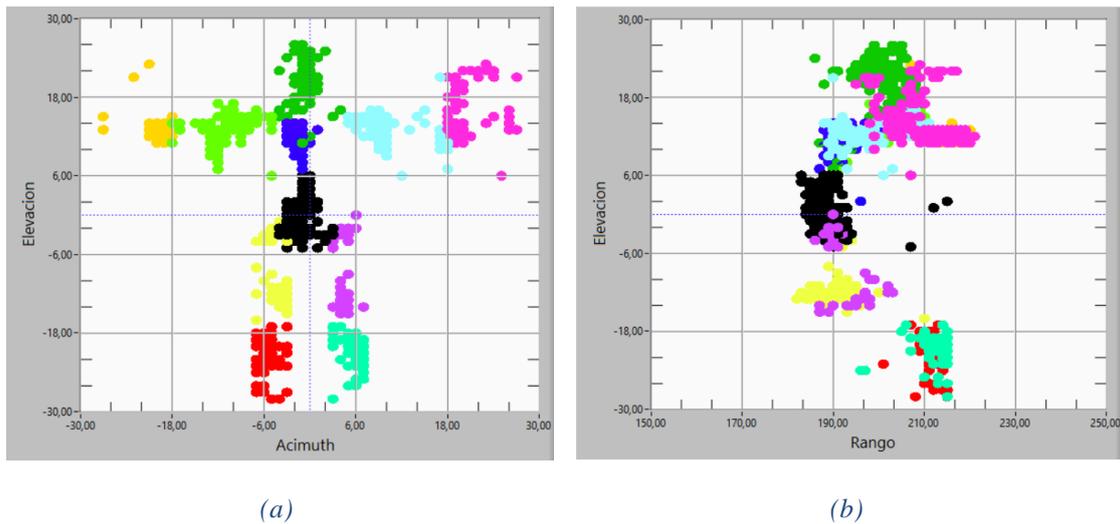
Tabla 5: Límites espaciales de los contenedores tridimensionales.  $\theta$  ( $^\circ$ ),  $\varphi$  ( $^\circ$ ) y  $r$  (cm)

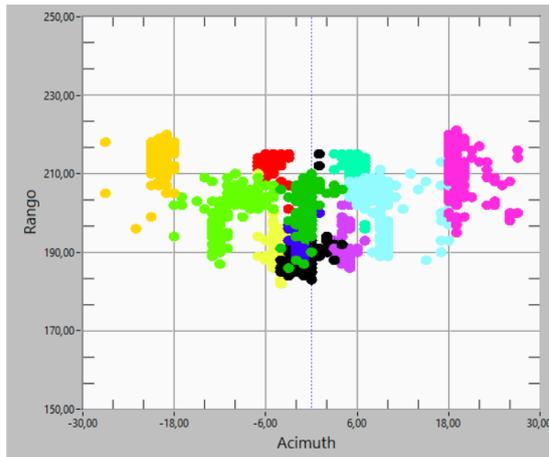
Los máximos son procesados por los contenedores obteniendo a la salida sólo los máximos que estaban contenidos en alguno de ellos. En caso de que no haya ninguno se deja vacío y en caso de que haya dos, se toma el que mayor amplitud presenta. Finalmente, según el contenedor al que se han asociado, se etiquetan de manera correspondiente para poder asociar el máximo con la parte del cuerpo que representa.

Los individuos que utilicen el sistema pueden presentar figuras antropométricas muy variables entre sí haciendo imposible establecer unos parámetros para los contenedores que sean universales. Es decir, puede que donde una persona tenga definido su contenedor para la cabeza, para otra persona ese contenedor este asociado al pecho.

La manera de solucionar este problema ha sido considerar dos posiciones clave a partir de los cuáles se construyen el resto. En base a la gran información antropométrica que aportan y a su notable presencia en la gran mayoría de las imágenes [10], se han escogido como posiciones clave la de la cintura y la cabeza. Así se establecen los parámetros para el resto de los contenedores en función de la posición de los máximos asociados a cintura y cabeza del individuo. En caso de que alguno de estos blancos no esté presente en la imagen se utilizan unas posiciones clave estándar que permiten la detección coherente del resto de blancos.

A continuación, se muestra un ejemplo en el que se aplica el algoritmo de búsqueda inteligente para obtener a partir de 200 imágenes de un individuo (construidas con promedios de 10) los siguientes puntos característicos (*figura 62*).





(c)

Figura 62: Máximos característicos obtenidos a partir de 200 imágenes de un individuo. (a) Vista frontal, (b) Vista lateral, (c) Vista superior

A partir de la observación de los máximos obtenidos se pueden extraer varias conclusiones:

1. La dispersión de los datos es elevada por lo que el máximo que se hace corresponder a cabeza, pecho, etc., no se encuentra en una posición fija.
2. En muchas ocasiones no se encuentra un máximo en el espacio designado por el contenedor que corresponde al elemento de reflexión asociado. Es decir, los elementos del cuerpo escogidos no siempre producen una reflexión o esta no siempre se traduce en un máximo de energía.
3. Aunque se aprecia en algunas zonas de las imágenes cierta tendencia a presentar máximos de energía, la superficie de un individuo es compleja por lo que las reflexiones producidas difícilmente pueden ser caracterizadas como norma general por unas ubicaciones exactas del cuerpo humano. Es decir, los límites entre lo que se considera brazo y mano o parte superior e inferior de la pierna son difusos puesto que el máximo de energía no siempre se produce en un punto concreto.

## Capítulo 6. IDENTIFICACIÓN. SVM

Decidir la identidad de la persona a través de los parámetros extraídos de las imágenes es una tarea compleja. Dado que se considera la identificación de múltiples individuos, se trata de un problema de clasificación multi-clase. La situación planteada se resuelve a través de aprendizaje supervisado y mediante el uso de máquinas de vectores de soporte (SVM).

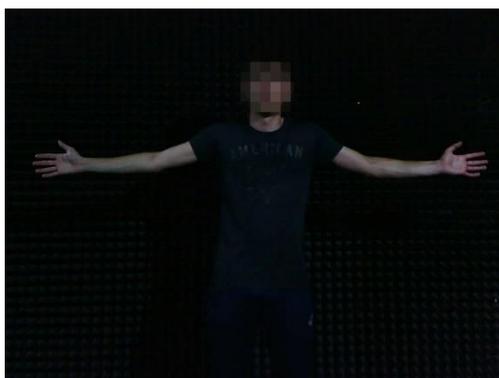
Para comprobar la viabilidad de este estudio, se ha construido un modelo biométrico a partir de 2000 imágenes capturadas de 3 personas diferentes (*tabla 5*).

	Clase:	Imágenes capturadas	Género	Altura
Individuo 1:	0	2000	Masculino	Alto
Individuo 2:	1	2000	Masculino	Media
Individuo 3:	2	2000	Femenino	Media

*Tabla 6: Características de las capturas acústicas*

Las imágenes acústicas se han capturado en tandas de 100. Típicamente y debido principalmente al estado de la red, pausas entre tandas y tiempos de procesado y guardado de las señales, capturar 2000 imágenes, es decir 20 tandas, supone un promedio de 30 minutos. De este modo, capturar las imágenes de los tres individuos propuestos conlleva aproximadamente 90 minutos.

Para comprobar la influencia de la vestimenta y verificar la robustez del algoritmo de clasificación, la mitad de las imágenes de los individuos 2 y 3 han sido capturadas con distinta ropa (*Figura 63: 12.a, 12.b, 13.a, 13.b*), mientras que la totalidad de las capturas realizadas sobre el individuo 1 se han tomado con la misma vestimenta (*Figura 63: 11*).



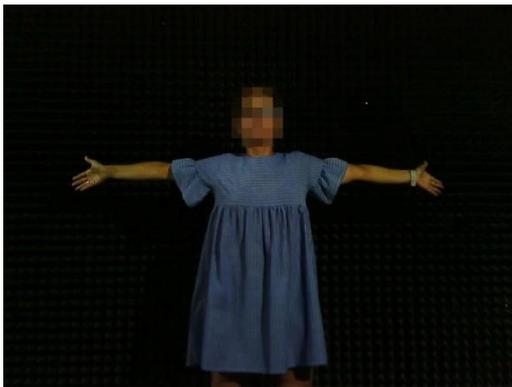
(11)



(12.a)



(12.b)



(13.a)



(13.b)

Figura 63: Individuo 11 e individuos 12 e 13 con sus respectivas vestimentas (a y b)

Como se ha comentado anteriormente, en la etapa de procesado se realiza un promediado de las imágenes con el fin de compensar micro movimientos de la persona capturada. Dependiendo de si este promedio se realiza utilizando grupos de 20 o de 10 imágenes, el número de muestras disponibles para la etapa de aprendizaje automático será de 100 o 200 muestras respectivamente.

Se han tomado las mismas capturas para cada individuo de forma que el problema de clasificación pueda partir de un conjunto de datos balanceado, es decir, con el mismo número de muestras de cada clase. Esto permite encontrar un modelo más preciso y evitar las dificultades de tratar con un conjunto de datos desbalanceado (*imbalanced data set*) que produzca un sesgo hacia la clase mayoritaria [40].

En LabVIEW, la paleta de rutinas “*Analytics and Machine Learning VIs*” permite crear aplicaciones de aprendizaje automático que descubran patrones y generen modelos que reconozcan esos patrones en nuevos datos [41]. En particular se utilizará la paleta de rutinas de clasificación “*Classification VIs*” que permite entrenar modelos utilizando un algoritmo SVM.

Por último, con relación a la estructura del presente capítulo, el proceso de aprendizaje automático se ha organizado en los apartados “Entrenamiento” y “Clasificación”. En el primero se recoge la elaboración del algoritmo mientras que en el segundo se evalúan los distintos clasificadores obtenidos y su desempeño.

## 6.1 ENTRENAMIENTO

En esta sección se trata la preparación de los datos para el sistema de aprendizaje automático y la configuración de parámetros y validación del mecanismo SVM utilizado. Todo esto permite construir distintos clasificadores cuyo desempeño es evaluado posteriormente en la sección de clasificación.

### 6.1.1 DEFINICIÓN Y EXTRACCIÓN DE CARACTERÍSTICAS

Las características son las variables individuales e independientes que miden las propiedades de cada clase. La elección de las características es una decisión importante a la hora de implementar cualquier modelo de aprendizaje automático y deben ser suficientemente representativas de la clase.

Así, según el modelo presentado en el apartado 4.3.2, se han definido los siguientes elementos de reflexión:

1. Cabeza
2. Cadera
3. Pecho
4. Brazo izquierdo
5. Brazo derecho
6. Mano izquierda
7. Mano derecha
8. Parte superior de la pierna izquierda
9. Parte superior de la pierna derecha
10. Parte inferior de la pierna izquierda
11. Parte inferior de la pierna derecha

A partir de cada uno de ellos, se utiliza una cuaterna de características, es decir: su valor de amplitud  $k$ , acimut  $\theta$ , elevación  $\varphi$  y rango  $r$ . De este modo se obtienen un total de 44 características para cada muestra tomada.

Para clasificar los parámetros extraídos por el modelo de optimización, es necesario una etapa extra en la que se utilizan los contenedores adaptativos para etiquetar los blancos y eliminar los que se encuentren fuera de rango.

#### ***Training set:***

Es imprescindible construir lo que se conoce como *training set*. El conjunto de datos se distribuye entre el conjunto de entrenamiento y el de test. El primero se emplea en la obtención del clasificador, mientras que el segundo sirve para evaluar su desempeño. Se ha elegido repartir el conjunto de datos en un 70% de datos de entrenamiento y 30% de datos de test con el fin de evitar los problemas de sobreajuste que se mencionan en el marco teórico, por lo que la distribución de los datos es la que se recoge a continuación (*tabla 7*).

Conjunto de datos para cada individuo			
Imágenes empleadas para realizar los promedios	Imágenes ya promediadas	Training set	Test set
N=10	200	140	60
N=20	100	70	30

Tabla 7: Distribución del conjunto de datos de entrenamiento y de test

Sin embargo, en función del número de imágenes promediadas (N) para generar una muestra, modelo y método de extracción de características utilizado, el número de características extraídas de las imágenes será mayor o menor.

Si se observan gráficamente las probabilidades de detección de cada elemento según el método utilizado (figura 64), se puede observar fácilmente como independientemente del modelo utilizado, la cadera y la cabeza son los elementos más fácilmente detectables lo que los convierte en los mejores puntos de referencia para los contenedores adaptativos. Además, se puede ver como las piernas (parte inferior y superior) y pecho, son en general los elementos más difíciles de detectar.

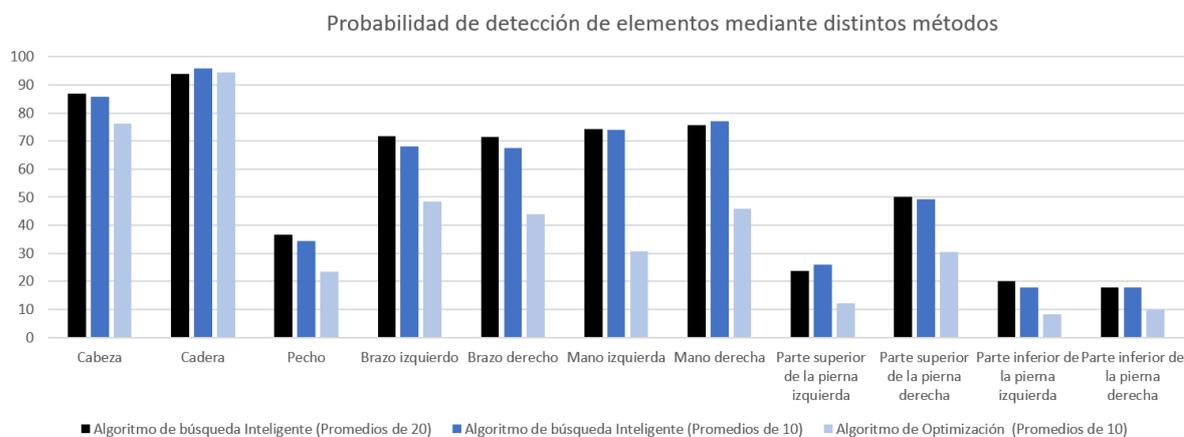


Figura 64: Probabilidades de detección de los elementos de reflexión mediante distintos métodos

Estos resultados denotan una gran variabilidad en la presencia de los blancos y un mejor desempeño del modelo inteligente puesto que dota al sistema de aprendizaje automático de un mayor número de características. Además, los resultados revelan que utilizar promedios de 20 imágenes no compensa en términos de detección de blancos puesto que los resultados son incluso ligeramente peores.

Para poder apreciar las pequeñas diferencias, en la tabla inferior se detallan las probabilidades de detección de cada cuaterna de características según el método utilizado (tabla 8).

Probabilidades de detección de cada elemento de reflexión (%)												
	1	2	3	4	5	6	7	8	9	10	11	Prom
<b>Modelo inteligente N=10</b>	85,67	95,83	34,50	68,17	67,50	73,83	77,00	25,83	49,17	17,83	17,83	56,54
<b>Modelo inteligente N=20</b>	87,00	94,00	36,67	71,67	71,33	74,33	75,67	23,67	50,00	20,00	17,67	55,74
<b>Modelo Optimización N=10</b>	76,33	94,50	23,33	48,33	43,83	30,83	45,83	12,33	30,33	8,33	9,83	38,52

Tabla 8: Probabilidades de detección de cada elemento de reflexión según el método y número de imágenes promediadas (N)

Si se comparan los modelos a alto nivel, es decir, calculando el promedio de detección de los 11 elementos, se obtiene una extracción del 56,54% y 55,74% de las características consideradas para el modelo inteligente utilizando promedios de 10 y 20 imágenes respectivamente y 38,52% para el modelo de optimización con promedios de 10 imágenes.

En las tablas a continuación (tablas 9 y 10), se ilustran algunas consideraciones del proceso de extracción de características para promedios de 10 imágenes a través de cada uno de los métodos empleados. El campo “Errores” recoge el porcentaje errores de convergencia que imposibilitan la extracción de características a partir del algoritmo de optimización por lo que este fenómeno no se da para el Modelo inteligente (tabla 10).

Optimización SQP + Identificación con modelo de contenedores. Promedios de 10 imágenes				
	Clase:	Muestras tomadas:	Tiempo de extracción:	Errores
Individuo 1:	<b>0</b>	<b>200</b>	<b>2h 55m</b>	<b>12%</b>
Individuo 2:	<b>1</b>	<b>200</b>	<b>2h</b>	<b>8,5%</b>
Individuo 3:	<b>2</b>	<b>200</b>	<b>2h 16m</b>	<b>10,5%</b>

Tabla 9: Resultados del proceso de extracción de características mediante el método de optimización

Modelo inteligente de máximos 3D. Promedios de 10 imágenes				
	Clase:	Muestras tomadas:	Tiempo de extracción:	Errores
Individuo 1:	<b>0</b>	<b>200</b>	<b>22seg</b>	<b>0%</b>
Individuo 2:	<b>1</b>	<b>200</b>	<b>20seg</b>	<b>0%</b>
Individuo 3:	<b>2</b>	<b>200</b>	<b>21seg</b>	<b>0%</b>

Tabla 10: Resultados del proceso de extracción de características mediante el modelo inteligente

Primeramente, la gran diferencia entre los tiempos de extracción de características (más de dos horas frente a menos de medio minuto) certifica al modelo inteligente como una mejor alternativa en términos de rapidez. Además, debido a errores de convergencia, con el método de optimización se pierde un gran porcentaje de los datos (de media el 10,3%) lo que también disminuye las características extraídas disminuyendo la calidad del método.

Por último, haciendo una pequeña discusión en términos de ECM, se han calculado los errores obtenidos para 200 estimaciones del individuo 1 entre la imagen real y las imágenes acústicas sintéticas construidas a partir de los elementos de reflexión devueltos en cada fase de la obtención de características mediante el método de optimización. Primero tras el algoritmo inteligente de inicialización (línea negra), después a partir de los parámetros optimizados (línea roja) y finalmente con los parámetros finales que se han considerado susceptibles de ser un elemento de reflexión a través de los contenedores adaptativos (línea verde) (figura 65).

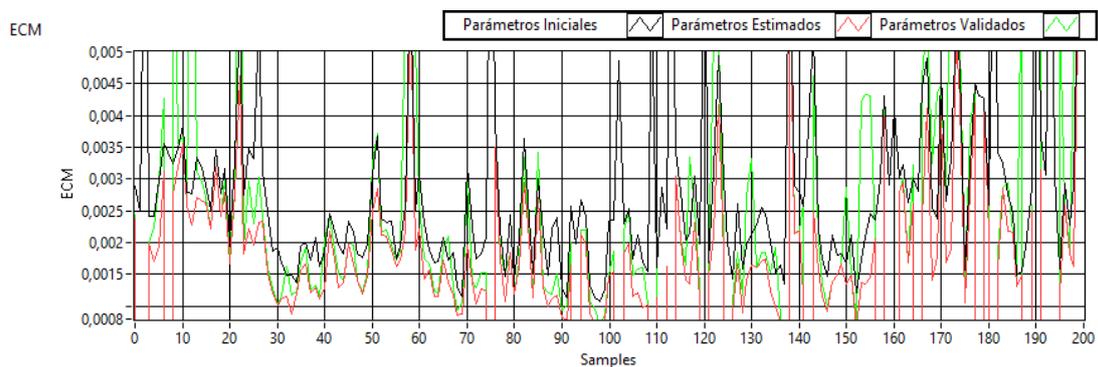


Figura 65: ECMs para las 200 muestras del individuo 1 con los parámetros iniciales, estimados y validados tras aplicar los contenedores.

Básicamente se puede ver como para todas las muestras, el ECM inicial (línea roja) se encuentra siempre por debajo del obtenido tras la optimización (línea negra) y tras los contenedores (línea verde), es decir el ECM es mayor con los parámetros iniciales, disminuye tras la estimación y vuelve a aumentar tras aplicar los contenedores adaptativos necesarios para clasificar los blancos (lo que se ha denominado como *fitting*)

Estos resultados evidencian que una buena aproximación a los parámetros no significa una imagen sintética más parecida a la real en términos de ECM.

### **Datos Incompletos:**

Como se ha visto, no todos los elementos de reflexión considerados se encuentran presentes en todas las capturas. Esto obliga a manejar datos incompletos. Es decir, casi en ninguna muestra se tiene información acerca de todos y cada uno de los elementos de reflexión determinados. Sin embargo, las características utilizadas para entrenar un algoritmo de *machine learning* a través de SVM deben de estar presentes en todas las imágenes lo que obliga a tratar los datos [42].

Una solución a este problema es eliminar las muestras o las características que contengan campos incompletos. Este es el enfoque más sencillo y aunque se pierde gran cantidad de

información genera un modelo robusto. Este planteamiento, debido a la gran cantidad de datos incompletos no se ha considerado para el caso de estudio particular.

Otra solución pasa por imputar los datos. Este es un problema aparte que muchas veces se resuelve propiamente con algoritmos de *machine learning* [42]. Un mecanismo de imputación simple es sustituir los valores que faltan utilizando un atributo cualitativo o cuantitativo de todos los valores no ausentes como por ejemplo la media o la moda, aunque también se pueden utilizar valores arbitrarios.

Es importante tener en cuenta la naturaleza de los datos incompletos. Para el caso de estudio, el hecho de que no se haya detectado algún elemento de reflexión también aporta información puesto que es una cualidad de la imagen y no un fallo del sistema. Es decir, los datos no faltan completamente al azar, sino que existe una dependencia con el individuo capturado. Para estos casos, también se puede agregar una variable adicional binaria que permita no perder información sobre la ausencia de los datos en caso de por ejemplo imputarlos a través de la media.

En resumen, para este proyecto, es lógico asumir 0 como el valor de amplitud de los blancos o máximos inexistentes. De este modo la amplitud recoge la cualidad de ausencia del blanco a considerar y los valores de acimut, elevación y rango se pueden sustituir por 0 como valor arbitrario.

Otra solución sería utilizar otros algoritmos de aprendizaje automático que soporten los datos incompletos, pero esto implicaría cambiar el enfoque de aprendizaje automático hacia otros algoritmos.

### 6.1.2 HIPERPARÁMETROS

La paleta de rutinas “*Analytics and Machine Learning VIs*” permite utilizar C-SVM o Nu-SVM. Ambos algoritmos son equivalentes, aunque el primero utiliza el parámetro “ $c$ ” mientras que el segundo utiliza la letra griega “ $\nu$ ” [41].

Previo al entrenamiento, la rutina de LabView encargada de generar el clasificador optimiza los hiperparámetros acorde con la métrica de evaluación especificada. Para ello se realiza una búsqueda exhaustiva a través de todas las posibles combinaciones. Se ha escogido “precisión” como métrica de evaluación, aunque los resultados obtenidos utilizando “exactitud” han sido similares. Ambas métricas se explicarán con detalle más adelante.

Los hiperparámetros considerados por la rutina son los siguientes [41], [43]:

- **Tipo de SVM (C-SVM o Nu-SVM):** Permite utilizar  $c$  o  $\nu$  como hiperparámetro.
- **$c$ :** Es el parámetro utilizado para el modelo C-SVM y debe ser mayor que 0. Los valores grandes de  $c$  aumentan la penalización por errores de clasificación obteniendo así márgenes más estrechos. Además, los valores grandes de  $c$  también corresponden a menos vectores de soporte puesto que hay menos observaciones mal clasificadas.
- **$\nu$ :** Es el parámetro utilizado para el modelo Nu-SVM y debe estar comprendido en el rango  $[0,1]$ .  $\nu$  es tanto un límite inferior para el número de muestras que son vectores de soporte como un límite superior para el número de muestras que están en el lado equivocado del hiperplano.

- **Tipo de función del kernel (lineal, polinómico, RBF, Sigmoide):** El kernel permite la transformación de los datos de entrada a otro espacio de mayor dimensión. De manera general, los kernel no lineales son más flexibles entre los que destaca como elección popular el kernel gaussiano o RBF. Debido a algunas propiedades teóricas indeseadas, el kernel sigmoide no resulta tan utilizado. En este documento no se justifican las expresiones matemáticas que construyen los distintos núcleos y permiten discriminar unas clases de otras pero se deja una referencia a disposición del lector para comprender los detalles [43].

Dependiendo del tipo de kernel utilizado, se utilizan o no los siguientes hyperparámetros:

- **Grado ( $d$ ):** especifica el coeficiente de grado de la función kernel polinómica e indica el grado del polinomio.
- **Gamma ( $\gamma$ ):** especifica el coeficiente gamma de la función kernel RBF, polinómica y sigmoide.
- **Coef0 ( $\beta_0$ ):** especifica el coeficiente coef0 de la función kernel polinómica y sigmoide.

### 6.1.3 VALIDACIÓN

La rutina emplea el método de validación cruzada explicado previamente en el marco teórico. Se especifican 3 pliegues por lo que el VI divide los datos de entrenamiento en 3 secciones.

Al tratarse de un problema de clasificación multiclase, para configurar la métrica de evaluación en la fase de validación se especifica una estrategia de promedio “micro”, es decir se calculan los valores numéricos de cada muestra y se devuelve la media de la métrica de todas las muestras. Este valor será el utilizado para medir el desempeño del clasificador.

## 6.2 CLASIFICACIÓN

Una vez construido el clasificador, se utilizan los datos reservados que forman parte del *test set* para evaluar el desempeño del algoritmo construido utilizando nuevas muestras.

En esta sección, primeramente, se presentan las principales métricas de evaluación que se pueden calcular a partir de la matriz de confusión y que permiten proporcionar una idea cuantitativa de la calidad del clasificador. Finalmente, y para concluir el capítulo, se exponen los resultados obtenidos a través de los diferentes modelos planteados en la memoria.

### 6.2.1 MÉTRICAS DE EVALUACIÓN:

Para evaluar la calidad de un algoritmo de aprendizaje automático, existen distintas métricas que permiten desentrañar su rendimiento según distintos aspectos. Estas métricas pueden ser de rango, umbral o probabilidad, aunque en la práctica las métricas de umbral suelen ser de las más utilizadas para medir el rendimiento de los clasificadores [44].

Para definir estas métricas, es preciso explicar antes lo que significan los términos  $tp$ ,  $tn$ ,  $fp$  y  $fn$  que se pueden calcular a partir de los datos obtenidos de la matriz de confusión explicada previamente en el marco teórico:

- **Verdadero positivo ( $tp$ ):** Total de casos en los que la muestra real se tomó sobre el individuo  $i$  y el modelo predice correctamente que la muestra correspondía a la de ese individuo.
- **Verdadero negativo ( $tn$ ):** Número de situaciones en las que la muestra real no se tomó sobre el individuo  $i$  y el modelo predice correctamente que la muestra no corresponde a dicho individuo.
- **Falso positivo ( $fp$ ):** Número de casos en los que la muestra real no se tomó sobre el individuo  $i$ , pero el modelo predice erróneamente que la muestra si corresponde al individuo  $i$ .
- **Falso negativo ( $fn$ ):** Cantidad de casos en los que la muestra real se tomó sobre el individuo  $i$ , pero el modelo predice erróneamente que la muestra no corresponde al individuo  $i$ .

**Exactitud (Accuracy):**

Tanto en problemas de clasificación binaria como en problemas de clasificación multiclase, la exactitud es la métrica de evaluación más utilizada para discriminar y seleccionar la solución óptima. Además, es una métrica sencilla y fácil de calcular [44].

La exactitud representa la relación entre las predicciones correctas sobre el total de predicciones evaluadas y se puede calcular para la clase  $i$  como:

$$Accuracy_i = \frac{tp_i + tn_i}{tp_i + fp_i + tn_i + fn_i} \tag{6.1}$$

Además, como una métrica más general que dé idea de cómo funciona el sistema, se puede calcular también la exactitud media a partir de la exactitud de cada clase [44].

**Tasa de error (Error rate):**

La tasa de error es la métrica complementaria de la exactitud, por lo que la suma de ambos debe ser igual a 1. Mientras que la exactitud mide la relación de las predicciones correctas, la tasa de error mide la proporción de predicciones incorrectas sobre el número total de predicciones evaluadas [44]. El parámetro se puede calcular para cada clase  $i$  como:

$$Error Rate_i = \frac{fp_i + fn_i}{tp_i + fp_i + tn_i + fn_i} \tag{6.2}$$

De nuevo, con la idea de obtener una métrica más general también se puede calcular la tasa de error media [44].

***Precisión (precision):***

Es importante no confundir la precisión con la exactitud. Mientras que la exactitud es una medida más genérica, la precisión se centra en las predicciones positivas. Es decir, la precisión mide las identificaciones positivas que se realizan correctamente sobre el total de predicciones positivas [44].

Traducido al contexto de identificación biométrica, mide el número de equivocaciones que se producen al decir que una persona es ella misma. Esta métrica por tanto es muy importante en el ámbito de estudio ya que mide las posibilidades de encontrarse con un impostor.

La precisión se calcula para cada clase  $i$  como:

$$Precision_i = \frac{tp_i}{tp_i + fp_i} \tag{6.3}$$

Al igual que los anteriores, también se puede calcular la precisión media como medida más general [44].

***Exhaustividad (recall):***

La exhaustividad, al igual que la precisión se centra en las predicciones positivas. Pero en este caso resuelve la pregunta de cuántos individuos pertenecientes a una clase fueron identificados correctamente respecto al total de individuos pertenecientes a esa clase [44].

No hay un convenio claro entre los significados de *recall* y *sensitivity* puesto que algunos estudios lo consideran lo mismo mientras que otros hacen diferencia entre los dos términos considerando los verdaderos positivos respecto al total de datos o sólo respecto al total de datos pertenecientes a esa clase. Acorde con el segundo criterio, el parámetro se puede modelar como se expone a continuación para la clase  $i$ :

$$Exhaustividad_i = \frac{tp_i}{tp_i + fn_i} \tag{6.4}$$

De nuevo, la exhaustividad media da una idea más general del desempeño del sistema [44].

***Especificidad (Specificity):***

Por otro lado, la especificidad se centra en las predicciones negativas. Aplicado al caso de estudio, la pregunta que la especificidad responde sería la siguiente: ¿Cuántos individuos que no son el sujeto  $i$  son identificados correctamente como que no son dicho sujeto?

Es importante que este sea un valor alto puesto que representa directamente la seguridad del sistema de identificación. Si el valor de la especificidad para un individuo  $i$  fuese 1, querría decir que ningún impostor podría hacerse pasar por dicho usuario [44]. Este valor se puede calcular a través de la siguiente fórmula:

$$Especificidad_i = \frac{tn_i}{tn_i + fp_i} \quad (6.5)$$

**Valor-F (F1 Score):**

Sería deseable que el modelo presentase unos valores de precisión y exhaustividad elevados, pero por desgracia estos valores normalmente se encuentran en tensión, es decir, mejorar la precisión típicamente reduce la exhaustividad y viceversa.

Ante esta problemática, surge el valor-F que representa la media armónica entre los valores de precisión y exhaustividad. Al igual que la exactitud, el valor-F, reporta buenos resultados como discriminador, aunque no se suele utilizar para problemas de clasificación multiclase [44]. El valor-F se puede calcular de la siguiente manera:

$$Valor - F_i = \frac{2 \cdot Precision_i \cdot Exhaustividad_i}{Precision_i + Exhaustividad_i} \quad (6.6)$$

**6.2.2 EXPOSICIÓN DE RESULTADOS**

A continuación, se presentan los resultados obtenidos para 4 casos diferentes, mostrando las matrices de confusión y métricas de evaluación obtenidas para cada una de las situaciones planteadas:

- Utilizando el modelo de optimización con valores absolutos y N=10.
- Utilizando el modelo inteligente de máximos 3D con valores normalizados y N=10
- Utilizando el modelo inteligente de máximos 3D con valores absolutos y N=10
- Utilizando el modelo inteligente de máximos 3D con valores absolutos y N=20

El objetivo es comprobar no sólo la influencia del modelo y el número de imágenes utilizadas al aplicar el promediado inicial sino también validar la importancia de los valores de amplitud absolutos como elemento diferenciador.

**Modelo de optimización (Valores absolutos, N=10)**

Empleando el modelo de optimización y utilizando promedios de 10 imágenes (N = 10), se obtiene la siguiente matriz de confusión para los 180 experimentos (tabla 11).

		Predicciones		
		ID 0	ID 1	ID 2
Valores reales	ID 0	50	7	3
	ID 1	4	51	5
	ID 2	5	16	39

*Tabla 11: Matriz de confusión generada utilizando el modelo de optimización con promedios de 10 imágenes*

A través de la matriz de confusión, se pueden calcular las métricas de evaluación expuestas en el apartado anterior para cada individuo y de manera global (tabla 12).

<b>Modelo de optimización (valores absolutos), N=10. Métricas de evaluación</b>						
	Exactitud	Tasa de Error	Precisión	Exhaustividad	Especificidad	Valor-F
<b>Individuo 0</b>	<b>0,894</b>	<b>0,106</b>	<b>0,847</b>	<b>0,833</b>	<b>0,925</b>	<b>0,840</b>
<b>Individuo 1</b>	<b>0,822</b>	<b>0,178</b>	<b>0,689</b>	<b>0,850</b>	<b>0,808</b>	<b>0,761</b>
<b>Individuo 2</b>	<b>0,839</b>	<b>0,161</b>	<b>0,829</b>	<b>0,650</b>	<b>0,933</b>	<b>0,729</b>
<b>Global</b>	<b>0,852</b>	<b>0,148</b>	<b>0,788</b>	<b>0,778</b>	<b>0,889</b>	<b>0,777</b>

Tabla 12: Métricas de evaluación obtenidas utilizando el modelo de optimización con promedios de 10 imágenes

### **Modelo Inteligente de máximos 3D (Valores Normalizados, N=10)**

Con el fin de comprobar la relevancia de los valores absolutos de amplitud de los máximos, se ha entrenado un modelo con promedios de 10 imágenes en el que todos los valores de amplitud permanecen normalizados por un factor igual al máximo absoluto de cada imagen.

Bajo estas circunstancias, se obtiene la matriz de confusión para los 180 experimentos que se muestra a continuación (tabla 13).

		<b>Predicciones</b>		
		<b>ID 0</b>	<b>ID 1</b>	<b>ID 2</b>
<b>Valores reales</b>	<b>ID 0</b>	51	5	4
	<b>ID 1</b>	5	49	6
	<b>ID 2</b>	5	17	38

Tabla 13: Matriz de confusión generada utilizando el modelo inteligente con promedios de 10 imágenes y valores de amplitud normalizada

De manera análoga al caso anterior, se presentan también las métricas de evaluación correspondientes (tabla 14).

<b>Modelo inteligente de máximos 3D Normalizados, N=10. Métricas de evaluación</b>						
	Exactitud	Tasa de Error	Precisión	Exhaustividad	Especificidad	Valor-F
<b>Individuo 0</b>	<b>0,894</b>	<b>0,106</b>	<b>0,836</b>	<b>0,85</b>	<b>0,917</b>	<b>0,843</b>
<b>Individuo 1</b>	<b>0,817</b>	<b>0,183</b>	<b>0,690</b>	<b>0,817</b>	<b>0,817</b>	<b>0,748</b>
<b>Individuo 2</b>	<b>0,822</b>	<b>0,178</b>	<b>0,792</b>	<b>0,633</b>	<b>0,917</b>	<b>0,704</b>
<b>Global</b>	<b>0,844</b>	<b>0,156</b>	<b>0,773</b>	<b>0,767</b>	<b>0,884</b>	<b>0,765</b>

Tabla 14: Métricas de evaluación obtenidas utilizando el modelo inteligente con promedios de 10 imágenes y valores de amplitud normalizada

**Modelo Inteligente de máximos 3D (Valores Absolutos, N=10)**

A partir de promedios de 10 imágenes aplicando el modelo inteligente de máximos 3D se obtiene la matriz de confusión para los 180 experimentos que se ilustra seguidamente (tabla 15) y calculando las métricas de evaluación como habitualmente (tabla 16).

		Predicciones		
		ID 0	ID 1	ID 2
Valores reales	ID 0	55	4	1
	ID 1	3	49	8
	ID 2	1	9	50

Tabla 15: Matriz de confusión generada utilizando el modelo inteligente con promedios de 10 imágenes y valores absolutos de amplitud

Modelo inteligente de máximos 3D absolutos, N=10. Métricas de evaluación						
	Exactitud	Tasa de Error	Precisión	Exhaustividad	Especificidad	Valor-F
<b>Individuo 0</b>	<b>0,95</b>	<b>0,05</b>	<b>0,932</b>	<b>0,917</b>	<b>0,967</b>	<b>0,924</b>
<b>Individuo 1</b>	<b>0,867</b>	<b>0,133</b>	<b>0,790</b>	<b>0,817</b>	<b>0,892</b>	<b>0,803</b>
<b>Individuo 2</b>	<b>0,895</b>	<b>0,106</b>	<b>0,847</b>	<b>0,833</b>	<b>0,925</b>	<b>0,840</b>
<b>Global</b>	<b>0,904</b>	<b>0,096</b>	<b>0,856</b>	<b>0,856</b>	<b>0,928</b>	<b>0,856</b>

Tabla 16: Métricas de evaluación obtenidas utilizando el modelo inteligente con promedios de 10 imágenes y valores absolutos de amplitud

**Modelo Inteligente de máximos 3D (Valores Absolutos, N=20)**

Por último, se evalúa también el modelo inteligente de máximos 3D, pero en esta ocasión utilizando promedios de 20 imágenes, que decrecen el número de experimentos a 90 reduciéndose la fiabilidad. Así se obtiene la siguiente matriz de confusión y métricas de evaluación (tablas 17 y 18).

		Predicciones		
		ID 0	ID 1	ID 2
Valores reales	ID 0	29	1	0
	ID 1	2	24	4
	ID 2	2	6	22

Tabla 17: Matriz de confusión generada utilizando el modelo inteligente con promedios de 20 imágenes y valores absolutos de amplitud

<b>Modelo inteligente de máximos 3D absolutos, N=20. Métricas de evaluación</b>						
	Exactitud	Tasa de Error	Precisión	Exhaustividad	Especificidad	Valor-F
<b>Individuo 0</b>	<b>0,944</b>	<b>0,056</b>	<b>0,879</b>	<b>0,967</b>	<b>0,933</b>	<b>0,921</b>
<b>Individuo 1</b>	<b>0,856</b>	<b>0,144</b>	<b>0,774</b>	<b>0,8</b>	<b>0,883</b>	<b>0,787</b>
<b>Individuo 2</b>	<b>0,867</b>	<b>0,133</b>	<b>0,846</b>	<b>0,733</b>	<b>0,933</b>	<b>0,786</b>
<b>Global</b>	<b>0,889</b>	<b>0,111</b>	<b>0,833</b>	<b>0,833</b>	<b>0,916</b>	<b>0,831</b>

Tabla 18: Métricas de evaluación obtenidas utilizando el modelo inteligente con promedios de 20 imágenes y valores absolutos de amplitud

### 6.2.3 ANÁLISIS DE RESULTADOS

A continuación, en base a las métricas de evaluación globales, se compara gráficamente la calidad de los métodos planteados (figura 66).

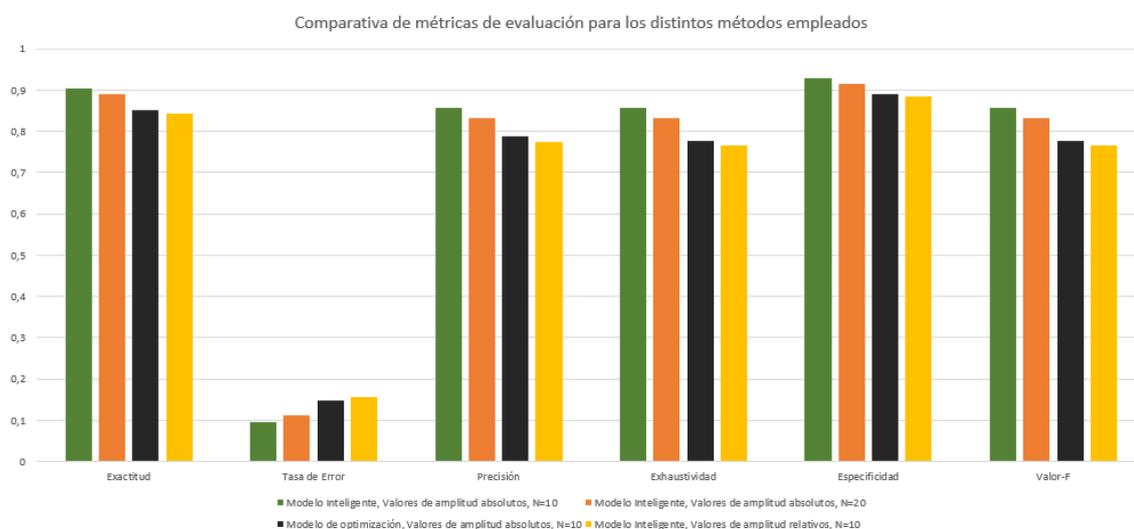


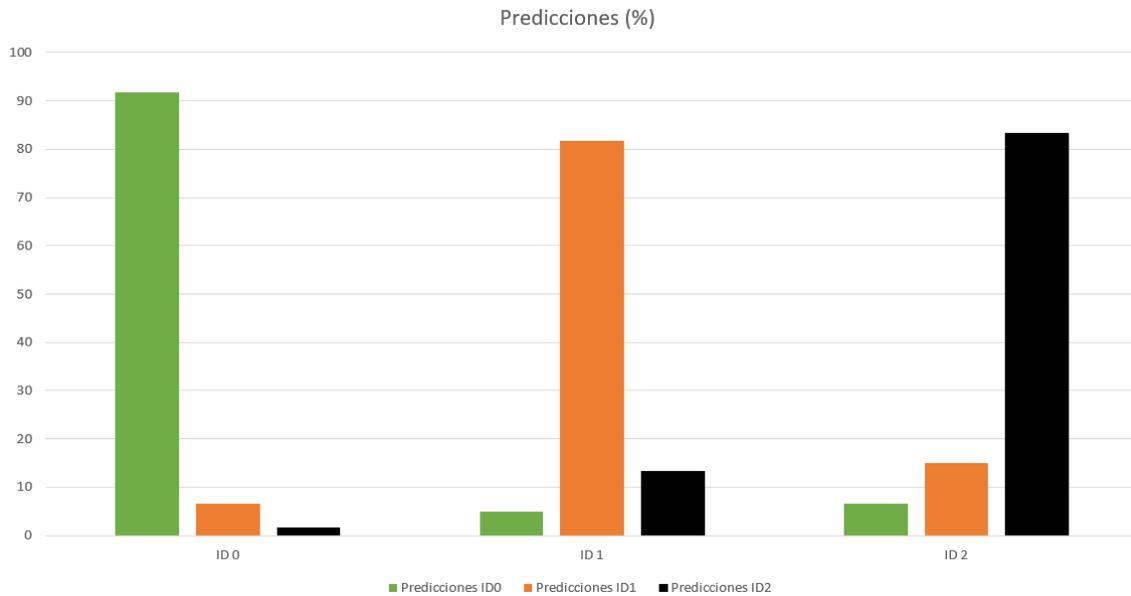
Figura 66: Comparativa de las métricas de evaluación globales para cada método aplicado

De este modo, los mejores resultados se han obtenido utilizando el modelo inteligente de máximos 3D con valores de amplitud absolutos y utilizando promedios de 10 imágenes (verde), mientras que los peores resultados se han obtenido a partir del modelo que utiliza los valores de amplitud relativos (amarillo) seguido del modelo de optimización (negro).

En relación con el uso de mayor número de imágenes a la hora de realizar los promedios, la conclusión es que además de suponer una disminución en la cantidad de datos disponibles, el desempeño del clasificador obtenido no mejora utilizando promedios de 20 imágenes por lo que utilizar promedios de 10 imágenes es la mejor opción.

Por otra parte, aprovechar los valores absolutos de amplitud contenidos en las imágenes permite obtener un mejor clasificador respecto al construido a partir de los valores normalizados.

Analizando individualmente los resultados proporcionados por el mejor modelo, se ha construido una gráfica donde para el conjunto de experimentos de cada individuo (ID0, ID1, ID2), se representa el porcentaje de predicciones realizadas correcta o incorrectamente para cada uno de los tres sujetos posibles (figura 67).



*Figura 67: Predicciones (%) realizadas a través del clasificador construido con el modelo inteligente de máximos 3D, valores de amplitud absolutos y N=10*

Se puede ver como el número de errores cometidos por el sistema a la hora de predecir la identidad del individuo perteneciente a la clase 0 es menor que para las clases 1 y 2. Esto en parte se debe a que las capturas del primer individuo fueron tomadas con la misma vestimenta por lo que es razonable conseguir un mejor desempeño del clasificador.

El clasificador final cuenta por tanto con una especificidad de 0,928 y una precisión de 0,856 como medidas globales. Es importante que ambas medidas sean altas puesto que representan de manera directa la calidad del sistema de identificación biométrico.

## Capítulo 7. CONCLUSIONES

---

A la hora de construir un clasificador, el modelo inteligente de máximos 3D proporciona mejores resultados que el modelo basado en optimización. Pese a la capacidad del modelo complejo de construir imágenes sintéticas más similares a la imagen real, puesto que se basa en minimizar los ECMs obtenidos, no se obtienen unas características de mayor calidad.

Aplicar una estrategia de mínimos cuadrados es un buen planteamiento si el modelo matemático empleado representa fielmente la realidad. Se ha visto como los resultados a la hora de estimar la posición de unos balones separados mejoran cuando se utiliza un modelo más riguroso. No obstante, cuando las imágenes son tomadas a partir de personas reales, la hipótesis que permite considerar la imagen formada a partir de un conjunto finito de blancos puntuales deja de ser cierta puesto que, a diferencia de los balones, la figura humana es una superficie mucho más compleja.

El algoritmo de optimización debe partir de unos buenos parámetros iniciales para realizar una búsqueda que permita obtener los valores óptimos finales. Sin embargo, inicializar correctamente el modelo sin información a priori resulta una tarea complicada y puede incurrir en un error de convergencia. Esto requiere reinicializar el algoritmo con unos parámetros de inicialización modificados de forma aleatoria.

Además, al aumentar el orden del modelo (número de blancos considerados), aumentan las variables de optimización lo que sumado al hecho de que la función objetivo no sea lineal complican el proceso. Esto hace que la función objetivo cuente con un gran número de máximos locales y por ello aumente el tiempo de optimización y la dependencia del algoritmo con los parámetros iniciales.

Por otra parte, en relación con las imágenes acústicas, se ha aplicado con éxito una estrategia de promediado. Este procesado mitiga la variabilidad en las capturas causada por los micro movimientos y pequeños cambios de postura del individuo escaneado. Se ha demostrado también la influencia de la vestimenta de la persona, aunque el modelo presentado es considerablemente robusto para sobreponerse a esta dificultad.

Respecto al análisis efectuado mediante el algoritmo de clasificación SVM (capítulo 6), se trata de un estudio preliminar realizado sobre la base de 3 individuos que debe ser evaluado con un número significativo de personas.

Para finalizar, aunque los resultados obtenidos no son aún lo suficientemente precisos como para establecer un sistema de autenticación biométrica unimodal en sí mismo, fusionar la información obtenida con otros métodos biométricos, permitiría conseguir un sistema biométrico multimodal que mejore significativamente la precisión [18].

## 7.1 LÍNEAS FUTURAS

La escasa investigación existente en la literatura para la identificación biométrica mediante imágenes acústicas, unido al trabajo realizado donde una gran cantidad de elementos se han tenido en cuenta en el diseño del clasificador, permite entrever las numerosas oportunidades de desarrollo e investigación en esta línea.

Mejorar el sistema de adquisición es sencillo incrementando el número de sensores o modificando el espaciado entre ellos. Esto permite contar con mayor resolución, y aunque permita resolver problemas como la detección de objetos próximos como dos blancos independientes, no se garantiza que se obtengan mejores resultados biométricos.

El uso de algoritmos heurísticos disminuye considerablemente la elevada carga computacional que implica un algoritmo de optimización, por lo que permite aumentar el volumen de datos a utilizar. De esta forma, se podría aumentar la resolución espacial del array acústico o utilizar por ejemplo la información multifrecuencial de las imágenes para mejorar la calidad de las características biométricas.

Dada la variabilidad en la presencia y posición de los máximos presentes en las imágenes de las personas, elaborar un modelo heurístico basado en la obtención de información acerca de distintas zonas de energía en lugar de puntos concretos del espacio puede resultar en un mejor identificador.

De esta manera, este trabajo ha sentado las bases para una nueva línea de investigación en el área de la clasificación biométrica mediante imágenes acústicas de personas.

## BIBLIOGRAFÍA

---

- [1] A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan. 2004, doi: 10.1109/TCSVT.2003.818349.
- [2] D. v. Helversen, M. W. Holderied, and O. v. Helversen, “Echoes of bat-pollinated bell-shaped flowers: conspicuous for nectar-feeding bats?,” *Journal of Experimental Biology*, vol. 206, no. 6, pp. 1025–1034, Mar. 2003, doi: 10.1242/JEB.00203.
- [3] Z. Song, K. Nguyen, T. Nguyen, C. Cho, and J. Gao, “Spartan Face Mask Detection and Facial Recognition System,” *Healthcare 2022, Vol. 10, Page 87*, vol. 10, no. 1, p. 87, Jan. 2022, doi: 10.3390/HEALTHCARE10010087.
- [4] Y. Guo, “Impact on Biometric Identification Systems of COVID-19,” *Sci Program*, vol. 2021, 2021, doi: 10.1155/2021/3225687.
- [5] B. Cafer, A. Emrullah, and Ç. Abidin, “Comparison of Multiple Biometric Identification with a Single Biometric Identification System | Request PDF,” in *Comparison of Multiple Biometric Identification with a Single Biometric Identification System*. , 2018. Accessed: Jul. 25, 2022. [Online]. Available: [https://www.researchgate.net/publication/328420862\\_Comparison\\_of\\_Multiple\\_Biometric\\_Identification\\_with\\_a\\_Single\\_Biometric\\_Identification\\_System](https://www.researchgate.net/publication/328420862_Comparison_of_Multiple_Biometric_Identification_with_a_Single_Biometric_Identification_System)
- [6] A. Izquierdo Fuente, L. del Val, M. I. Jiménez, and J. J. Villacorta, “Performance evaluation of a biometric system based on acoustic images,” *Sensors*, vol. 11, no. 10, pp. 9499–9519, Oct. 2011, doi: 10.3390/S111009499.
- [7] S. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, “Data Preprocessing for Supervised Learning,” *Int J Comp Sci*, pp. 111–117, 2006, Accessed: Jul. 25, 2022. [Online]. Available: [https://www.researchgate.net/publication/228084519\\_Data\\_Preprocessing\\_for\\_Supervised\\_Learning](https://www.researchgate.net/publication/228084519_Data_Preprocessing_for_Supervised_Learning)
- [8] M. Moebus, A. M. Zoubir, and M. Viberg, “Parametrization of acoustic images for the detection of human presence by mobile platforms,” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 3538–3541, 2010, doi: 10.1109/ICASSP.2010.5495940.
- [9] M. Moebus and A. M. Zoubir, “Three-dimensional ultrasound imaging in air using a 2D array on a fixed platform,” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2, 2007, doi: 10.1109/ICASSP.2007.366397.
- [10] M. Moebus, “Object Detection and Classification for Mobile Platforms Using 3D Acoustic Imaging,” Technische Universität Darmstadt, Darmstadt, 2011.
- [11] N. Srivastava, U. Agrawal, S. K. Roy, and U. S. Tiwary, “Human identification using Linear Multiclass SVM and Eye Movement biometrics,” *2015 8th*

- 
- International Conference on Contemporary Computing, IC3 2015*, pp. 365–369, Dec. 2015, doi: 10.1109/IC3.2015.7346708.
- [12] D. Rezgui and Z. Lachiri, “ECG biometric recognition using SVM-based approach,” *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 11, pp. S94–S100, Jul. 2016, doi: 10.1002/TEE.22241.
- [13] S. Soviany, S. Puscoci, and V. Sandulescu, “A Biometric Identification System with Kernel SVM and Feature-level Fusion,” *Proceedings of the 12th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2020*, Jun. 2020, doi: 10.1109/ECAI50035.2020.9223188.
- [14] L. del Val, A. Izquierdo Fuente, J. J. Villacorta, and M. Raboso, “Acoustic Biometric System Based on Preprocessing Techniques and Linear Support Vector Machines,” *Sensors 2015, Vol. 15, Pages 14241-14260*, vol. 15, no. 6, pp. 14241–14260, Jun. 2015, doi: 10.3390/S150614241.
- [15] V. A. Manasson, R. M. Mino, and L. S. Sadovnik, “Spinning grating antenna for MMW imaging,” *Proceedings of the SPIE*, vol. 3064, pp. 134–143, Jun. 1997, doi: 10.1117/12.277074.
- [16] T. G. Nyland, J. S. Mattoon, E. J. Herrgesell, and E. R. Wisner, “Physical Principles, Instrumentation, and Safety of Diagnostic Ultrasound,” in *Small Animal Diagnostic Ultrasound*, 3rd ed., Saunders, 2015, p. 49. doi: 10.1016/B978-0-7216-7788-0.50007-1.
- [17] M. I. Skolnik, *Introduction to Radar Systems*, 1981st ed. Singapore: McGraw-Hill, 1981. Accessed: Jul. 28, 2022. [Online]. Available: [https://www.scirp.org/\(S\(i43dyn45teexjx455qlt3d2q\)\)/reference/ReferencesPapers.aspx?ReferenceID=64187](https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=64187)
- [18] R. Seemin and M. Ajaz H, “Biometrics Verification: a Literature Survey,” *Journal of Computing and ICT Research*, vol. 5, no. 2, pp. 67–80, Dec. 2011, Accessed: Jul. 28, 2022. [Online]. Available: [https://www.researchgate.net/publication/327667231\\_Biometrics\\_Verification\\_a\\_Literature\\_Survey](https://www.researchgate.net/publication/327667231_Biometrics_Verification_a_Literature_Survey)
- [19] J. Andress, *The basics of information security : understanding the fundamentals of InfoSec in theory and practice*, 2nd ed. Syngress, 2014.
- [20] A. Carmona Benjumea, *Aspectos antropométricos de la población laboral española aplicados al diseño industrial*. Madrid: Instituto Nacional de Seguridad e Higiene en el Trabajo. Ministerio de Trabajo e inmigración, 2003.
- [21] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, 1993rd ed., vol. I. Financial Times Prentice Hall, 1993. Accessed: Jul. 28, 2022. [Online]. Available: <http://www.amazon.com/dp/0133457117>
- [22] A. Vidal, “Algoritmos Heurísticos en Optimización,” Santiago de compostela, Jul. 2013.

- 
- [23] S. H. Zanakis and J. R. Evans, “Heuristic ‘Optimization’: Why, When, and How to Use It,” *Interfaces (Providence)*, vol. 11, no. 5, pp. 84–91, Oct. 1981, doi: 10.1287/INTE.11.5.84.
- [24] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 16th ed. Cambridge, United Kingdom: Cambridge University Press, 2014. doi: 10.1017/CBO9780511801389.
- [25] K. A. AlAfandy, H. Omara, M. Lazaar, and M. al Achhab, “Computational Intelligence and Applications for Pandemics and Healthcare,” *Medical Information Science Reference*, 2022, pp. 83–113. doi: 10.4018/978-1-7998-9831-3.CH005.
- [26] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 2010th ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [27] “Jeff Kodosky - National Instruments.” <https://www.ni.com/es-es/about-ni/leadership/kodosky.html> (accessed Aug. 01, 2022).
- [28] J. Kodosky, “LabVIEW,” *Proceedings of the ACM on Programming Languages*, vol. 4, no. HOPL, Jun. 2020, doi: 10.1145/3386328.
- [29] “Fundamentos del Entorno de LabVIEW - National Instruments.” <https://www.ni.com/getting-started/labview-basics/esa/environment> (accessed Aug. 01, 2022).
- [30] A. Izquierdo, J. J. Villacorta, L. del Val Puente, and L. Suárez, “Design and Evaluation of a Scalable and Reconfigurable Multi-Platform System for Acoustic Imaging,” *Sensors*, vol. 16, no. 10, p. 1688, Oct. 2016, doi: 10.3390/S16101671.
- [31] L. del Val, A. Izquierdo, J. J. Villacorta, and L. Suárez, “Using a Planar Array of MEMS Microphones to Obtain Acoustic Images of a Fan Matrix,” 2017, doi: 10.1155/2017/3209142.
- [32] STMicroelectronics, “MP34DT01 Datasheet.” Accessed: Aug. 23, 2022. [Online]. Available: [https://www.st.com/content/st\\_com/en.html](https://www.st.com/content/st_com/en.html)
- [33] “NI myRIO Hardware at a Glance - NI,” Feb. 04, 2020. <https://www.ni.com/es-es/support/documentation/supplemental/13/ni-myrio-hardware-at-a-glance.html> (accessed Aug. 02, 2022).
- [34] Y. Bai, L. Lu, J. Cheng, J. Liu, Y. Chen, and J. Yu, “Acoustic-based sensing and applications: A survey,” *Computer Networks*, vol. 181, p. 107470, Nov. 2020, doi: 10.1016/J.COMNET.2020.107447.
- [35] Rajiv Padhye and Rajkishore Nayak, *Acoustic Textiles*, 1st ed. Singapore: Springer Singapore, 2016. doi: 10.1007/978-981-10-1476-5.
- [36] G. S. K. Wong, “Speed of sound in standard air,” *Journal of the Acoustical Society of America*, vol. 79, no. 5, pp. 1359–1366, 1986, doi: 10.1121/1.393664.

- 
- [37] G. S. K. Wong and T. F. Embleton, “Variation of the speed of sound in air with humidity and temperature,” *Journal of the Acoustical Society of America*, vol. 77, no. 5, pp. 1710–1712, 1985, doi: 10.1121/1.391918.
- [38] P. T. Boggs and J. W. Tolle, “Sequential Quadratic Programming,” *Acta Numerica*, vol. 4, pp. 1–51, 1995, doi: 10.1017/S0962492900002518.
- [39] “Constrained Nonlinear Optimization VI - NI,” Apr. 30, 2022. [https://www.ni.com/docs/en-US/bundle/labview-2020/page/gmath/constrained\\_nonlinear\\_optimization.html](https://www.ni.com/docs/en-US/bundle/labview-2020/page/gmath/constrained_nonlinear_optimization.html) (accessed Aug. 06, 2022).
- [40] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*, 1st ed. Springer International Publishing, 2018. doi: 10.1007/978-3-319-98074-4.
- [41] “Analytics and Machine Learning VIs - NI,” Apr. 30, 2022. [https://www.ni.com/docs/en-US/bundle/labview-2018-analytics-and-machine-learning-toolkit-api-ref/page/lvaml/aml\\_pal.html](https://www.ni.com/docs/en-US/bundle/labview-2018-analytics-and-machine-learning-toolkit-api-ref/page/lvaml/aml_pal.html) (accessed Aug. 07, 2022).
- [42] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, “A survey on missing data in machine learning,” *J Big Data*, vol. 8, no. 1, pp. 1–37, Oct. 2021, doi: 10.1186/S40537-021-00516-9.
- [43] N. Guenther and M. Schonlau, “Support Vector Machines,” *Stata J*, vol. 16, no. 4, pp. 917–937, Dec. 2016, doi: 10.1177/1536867X1601600407.
- [44] H. M and S. M.N, “A Review on Evaluation Metrics for Data Classification Evaluations,” *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/IJDKP.2015.5201.