



Patrones Espacio–Acústicos en Sonidos Respiratorios Multicanal:
Integración de Procrustes, Clustering y Modelos Supervisados en
Pacientes con EPOC

Autor:

Sebastián Escobar Pajoy

Trabajo de grado

Asesores:

Diego Fernando Fonseca Valero, Profesor.

Escuela de Ciencias Aplicadas e Ingeniería.

Universidad EAFIT

UNIVERSIDAD EAFIT

ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2025

CONTENIDO

LISTA DE FIGURAS	4
LISTA DE TABLAS	5
RESUMEN	6
ABSTRACT	7
INTRODUCCIÓN	8
PLANTEAMIENTO DEL PROBLEMA.....	9
JUSTIFICACIÓN	10
OBJETIVOS.....	11
GENERAL	11
ESPECÍFICOS	11
MARCO CONCEPTUAL	12
ANÁLISIS Y CLASIFICACIÓN DE SONIDOS RESPIRATORIOS.....	12
AUSCULTACIÓN Y SONIDOS RESPIRATORIOS	14
DESCRIPTORES DE AUDIO PARA LA CARACTERIZACIÓN ACÚSTICA	21
REDUCCIÓN DE DIMENSIONALIDAD CON ESCALAMIENTO MULTIDIMENSIONAL CLÁSICO (MDS) Y ANÁLISIS DE PROCRUSTES.....	24
MODELOS SUPERVISADOS	27
DISEÑO METODOLÓGICO	29
1. ENTENDIMIENTO DEL PROBLEMA.....	29
2. ENTENDIMIENTO DE LOS DATOS	30
3. PREPARACIÓN DE LOS DATOS	32
4. MODELADO DE DATOS.....	36
5. EVALUACIÓN	43

RESULTADOS Y DISCUSIÓN	44
EXTRACCIÓN DE CARACTERISTICAS	47
ALINEAMIENTO GEOMETRICO	50
MODELADO SUPERVISADO	60
INTEGRACIÓN DEL ANÁLISIS SUPERVISADO Y NO SUPERVISADO	64
CONCLUSIONES	68
REFERENCIAS	70
ANEXOS.....	75
A. ORIGEN DE LOS DATOS	75
B. SCRIPTS UTILIZADOS	75

LISTA DE FIGURAS

<i>Figura 1: Grafica de onda seno para mostrar amplitud y frecuencia.....</i>	<i>15</i>
<i>Figura 2: Espectro de frecuencia</i>	<i>15</i>
<i>Figura 3: Ejemplo de alineamiento Procrustes</i>	<i>26</i>
<i>Figura 4: Ejemplo de segmentación de una señal de audio</i>	<i>33</i>
<i>Figura 5: Extracción de descriptor acústico en 3 bandas espectrales</i>	<i>34</i>
<i>Figura 6: Proceso seguido para la extracción y la transformación de datos</i>	<i>36</i>
<i>Figura 7: Proceso de algoritmo de alineamiento a referencia centroide</i>	<i>39</i>
<i>Figura 8: Proceso de estimación de k referencias centroides.....</i>	<i>41</i>
<i>Figura 9: Histogramas de duración de eventos acústicos wheeze</i>	<i>45</i>
<i>Figura 10: Histograma de duración de eventos acústicos crackle.....</i>	<i>45</i>
<i>Figura 11: Distribución de eventos por región torácica.....</i>	<i>47</i>
<i>Figura 12: Espectrograma de 2 canales respiratorios en donde ocurren crackles</i>	<i>48</i>
<i>Figura 13: Proyección con PCA de las en un espacio lineal</i>	<i>48</i>
<i>Figura 14: Mapa de calor de correlación de variables</i>	<i>49</i>
<i>Figura 15: Escalamiento multidimensional sobre el conjunto de 74368 segmentos individuales.....</i>	<i>50</i>
<i>Figura 16: Distribución del Stress normalizado para distancia logarítmica absoluta</i>	<i>51</i>
<i>Figura 17: Distribución del Stress normalizado para distancia euclídeana</i>	<i>51</i>
<i>Figura 18: Segmento multicanal procesado por escalamiento dimensional.....</i>	<i>52</i>
<i>Figura 19: Alineamiento promedio de todo el conjunto de datos, configuraciones con $D < 0.25$</i>	<i>53</i>
<i>Figura 20: Alineamiento promedio de todo el conjunto de datos, configuraciones con $D > 0.25$.....</i>	<i>53</i>
<i>Figura 21: Grafica del codo para inercia promedio</i>	<i>54</i>
<i>Figura 22: Muestras de clúster 0 alineadas.....</i>	<i>55</i>
<i>Figura 23: Muestras de clúster 1 alineadas.....</i>	<i>56</i>
<i>Figura 24: Muestras de clúster 2 alineadas</i>	<i>56</i>
<i>Figura 25: Muestras de clúster 3 alineadas.....</i>	<i>57</i>
<i>Figura 26: Muestras de clúster 4 alineadas.....</i>	<i>57</i>
<i>Figura 27: Muestras de clúster 5 alineadas.....</i>	<i>58</i>
<i>Figura 28: Distribución de eventos por clúster</i>	<i>59</i>
<i>Figura 29: Distribución de eventos por clúster y región torácica</i>	<i>60</i>
<i>Figura 30: Importancia de características tipo descriptor_canal</i>	<i>63</i>

LISTA DE TABLAS

<i>Tabla 1: Descripción de los datos</i>	<i>30</i>
<i>Tabla 2: Descriptores para cada banda de frecuencia</i>	<i>34</i>
<i>Tabla 3: Descriptores MFCC</i>	<i>34</i>
<i>Tabla 4: Conteo de eventos por región torácica.....</i>	<i>46</i>
<i>Tabla 5: Tamaño y compactación de cada clúster</i>	<i>58</i>
<i>Tabla 6: Métricas de desempeño para el modelo global</i>	<i>61</i>
<i>Tabla 7: Importancia por cada región torácica para el modelo global</i>	<i>62</i>
<i>Tabla 8: Top 10 de importancia promedio de descriptores para el modelo global</i>	<i>62</i>
<i>Tabla 9: Diferencias en rendimiento para pruebas de ablación</i>	<i>63</i>
<i>Tabla 10: Rendimiento del modelo global en cada clúster.....</i>	<i>65</i>
<i>Tabla 11: Características más importantes para la predicción de wheezes en el modelo global por clúster ...</i>	<i>65</i>
<i>Tabla 12: Rendimiento de los modelos entrenados por cada clúster.....</i>	<i>66</i>
<i>Tabla 13: Top 5 características para cada modelo de cada clúster</i>	<i>67</i>

RESUMEN

La auscultación pulmonar es una herramienta esencial en el diagnóstico de enfermedades respiratorias, pero su interpretación se ve limitada por la subjetividad del examinador y por la influencia de la posición de registro sobre las características acústicas. En este trabajo se propone una estrategia de análisis multicanal orientada a caracterizar la organización espacial de los sonidos respiratorios en pacientes con EPOC y a evaluar la contribución relativa de distintas regiones torácicas en la detección de eventos adventicios. A partir de grabaciones simultáneas en siete posiciones del tórax de la base ICBHI 2017, se construyó un conjunto de segmentos respiratorios descritos mediante características espectrales y cepstrales. Las configuraciones multicanal se proyectaron a espacios de baja dimensión y se alinearon mediante análisis de Procrustes, lo que permitió obtener representaciones geométricas comparables entre sujetos y respiraciones. Sobre estas representaciones se aplicó un esquema de agrupamiento no supervisado capaz de revelar patrones espaciales recurrentes asociados a diferentes distribuciones de sibilancias y crepitantes. Complementariamente, se entrenaron modelos supervisados basados en bosques aleatorios para la detección de eventos adventicios, incorporando análisis de importancia de características y experimentos de eliminación por canal para estudiar el aporte de cada región torácica. Los resultados muestran que la información espacial multicanal contiene patrones estructurados que pueden explotarse tanto para agrupar configuraciones torácicas como para mejorar la interpretabilidad de los modelos de clasificación, avanzando hacia representaciones más robustas y explicables de la acústica pulmonar.

Palabras clave: Auscultación pulmonar; EPOC; sonidos respiratorios; análisis multicanal; alineamiento Procrustes; escalamiento multidimensional; clustering; modelos supervisados; interpretabilidad.

ABSTRACT

Auscultation is a fundamental tool for assessing respiratory conditions; however, its interpretation is limited by examiner subjectivity and by the strong influence that recording location exerts on the acoustic properties of lung sounds. This work proposes a multichannel analysis strategy aimed at characterizing the spatial organization of respiratory sounds in patients with COPD and evaluating the relative contribution of different thoracic regions to the detection of adventitious events. Using simultaneous recordings from seven chest locations in the ICBHI 2017 dataset, a collection of respiratory segments was constructed and described through spectral and cepstral features. Multichannel configurations were projected into low-dimensional spaces and aligned using Procrustes analysis, enabling comparable geometric representations across subjects and breathing cycles. An unsupervised clustering scheme applied to these representations revealed recurrent spatial patterns associated with different distributions of wheezes and crackles. In addition, supervised Random Forest models were trained for adventitious sound detection, incorporating feature-importance analyses and channel-ablation experiments to examine the contribution of each thoracic region. The results indicate that multichannel spatial information contains structured patterns that can be leveraged both to group thoracic configurations and to enhance the interpretability of classification models, contributing to more robust and explainable representations of pulmonary acoustics.

Keywords: Lung auscultation; COPD; respiratory sounds; multichannel analysis; Procrustes alignment; multidimensional scaling; clustering; supervised learning; interpretability..

INTRODUCCIÓN

La auscultación pulmonar es un procedimiento clínico no invasivo fundamental para la evaluación del sistema respiratorio. Mediante la escucha de sonidos captados con un estetoscopio en regiones específicas del tórax, los profesionales pueden identificar patrones acústicos asociados con alteraciones fisiopatológicas [8][19]. Sonidos adventicios como sibilancias y crepitantes presentan correlaciones bien establecidas con enfermedades como neumonía, asma y enfermedad pulmonar obstructiva crónica (EPOC) [25].

La exploración se realiza sobre zonas supraclaviculares, infraclaviculares, mamarias, axilares y escapulares [14], lo que permite comparar regiones torácicas y detectar anomalías localizadas, como sonidos adventicios basales en procesos infecciosos o restrictivos. Estas diferencias regionales aportan información diagnóstica clave sobre la extensión y gravedad de las afecciones pulmonares [19].

Los avances recientes en análisis computacional han permitido abordar limitaciones de la auscultación tradicional, particularmente su dependencia del oído y experiencia del profesional. Métodos de procesamiento espectral, como transformadas de Fourier, wavelet y Hilbert-Huang, se han aplicado para extraer características relevantes de señales respiratorias [4][18]. En paralelo, modelos supervisados y no supervisados —incluyendo SVM, Random Forest y redes neuronales convolucionales— han mostrado capacidad para discriminar eventos acústicos en diversos entornos clínicos [9][16].

Pese a estos avances, persisten dificultades relacionadas con la variabilidad de los datos, la superposición espectral entre distintos tipos de sonidos y la limitada interpretabilidad de los modelos [27]. Estos desafíos han motivado el desarrollo de enfoques que buscan integrar información acústica, espacial y fisiológica, abriendo nuevas oportunidades para mejorar la comprensión y caracterización de los sonidos pulmonares.

PLANTEAMIENTO DEL PROBLEMA

La auscultación pulmonar, aunque esencial, presenta limitaciones inherentes a su naturaleza subjetiva. La identificación de sonidos respiratorios depende ampliamente de la experiencia del profesional, lo que introduce variabilidad interobservador significativa. Estudios previos han reportado niveles de concordancia entre 55 % y 80 % en la clasificación de sonidos respiratorios, evidenciando inconsistencias que afectan la precisión diagnóstica y la reproducibilidad clínica [17].

El análisis computacional ha intentado mitigar esta variabilidad mediante la extracción de características espectrales y el uso de modelos de aprendizaje automático. Sin embargo, persisten obstáculos importantes relacionados con la capacidad de generalización y la interpretabilidad de estos modelos. La limitada diversidad de las bases de datos disponibles y la superposición espectral entre distintos sonidos adventicios dificultan la discriminación automática de sibilancias y crepitantes [19].

Un vacío crítico, y aún poco abordado, es la influencia de la localización de grabación sobre las propiedades acústicas registradas. La evidencia demuestra variaciones sistemáticas según el sitio torácico y la postura del paciente: en sedestación, los sonidos inspiratorios presentan mayor intensidad en el hemitórax posterior izquierdo, mientras que en decúbito el pulmón dependiente registra mayor energía [35]. A pesar de las recomendaciones de estandarización de CORSA [25], la documentación de la posición de auscultación sigue siendo heterogénea. Esto compromete la reproducibilidad y puede inducir sesgos en los modelos computacionales, que pueden aprender patrones de sitio en lugar de patrones patológicos reales. Los registros multicanal ofrecen ventajas para estudiar la distribución espacial de los sonidos respiratorios, pero también introducen desafíos relacionados con la sincronización, la consistencia posicional y la anotación de cada canal [36].

Frente a este escenario, surge la necesidad de enfoques que no solo consideren la información espectral de cada canal, sino que integren explícitamente la dimensión espacial de los sonidos respiratorios. Representaciones capaces de capturar configuraciones y relaciones entre distintos puntos de auscultación permitirían identificar patrones recurrentes en la distribución regional de los sonidos adventicios, aportando una visión más robusta que los análisis basados únicamente en características locales. En este sentido, las técnicas de alineamiento geométrico aplicadas a arreglos multicanal han demostrado su utilidad para revelar relaciones entre sensores, como se ha observado en la correlación de señales EEG con

configuraciones de electrodos [36] y en la identificación de relaciones espaciales en sistemas acústicos multicanal [37].

Inspirado en estos enfoques, este proyecto explora estrategias que combinan la caracterización espectral con herramientas de análisis geométrico para detectar patrones espaciales en registros multicanal, además de incorporar elementos de interpretabilidad que permitan comprender en qué regiones del tórax y en qué tipos de sonidos se manifiestan con mayor intensidad las alteraciones respiratorias. De este modo, se busca avanzar hacia representaciones más robustas e informativas que mitiguen la variabilidad introducida por la posición de grabación y fortalezcan la utilidad clínica del análisis computacional de sonidos pulmonares.

JUSTIFICACIÓN

Tener herramientas robustas con resultados estables para el análisis de sonidos respiratorios puede influir en la estandarización de procesos y la toma de decisiones terapéuticas basadas en evidencia [22]. En este contexto, herramientas computacionales que permitan caracterizar los sonidos respiratorios de manera objetiva tienen un gran potencial como complemento de la práctica médica.

Sin embargo, la utilidad clínica de estos métodos depende de su capacidad para ser robustos frente a la variabilidad posicional y, simultáneamente, interpretables para los profesionales de la salud. La evidencia indica que diferencias sistemáticas entre regiones torácicas pueden inducir sesgos en los modelos, afectando su capacidad de generalizar y limitando su uso clínico [25] [36]. Por esta razón, resulta imprescindible desarrollar enfoques que incorporen explícitamente la dimensión espacial de los sonidos respiratorios.

Los arreglos multicanal abren la posibilidad de estudiar la distribución espacial de los sonidos y de construir representaciones más completas del fenómeno acústico [35]. La literatura muestra que el análisis geométrico y el alineamiento de configuraciones multicanal han sido útiles en neurofisiología y acústica para identificar patrones espaciales [36] [37]. Aplicar estas técnicas al ámbito respiratorio permite explorar relaciones entre regiones torácicas y reducir sesgos asociados a la posición de grabación.

Por tanto, este estudio se justifica en la necesidad de producir representaciones espaciales, robustas e interpretables de los sonidos pulmonares, capaces de

mejorar la comprensión clínica y de sentar bases más sólidas para futuras aplicaciones de aprendizaje automático en el diagnóstico de enfermedades respiratorias.

OBJETIVOS

GENERAL

Implementar estrategia de análisis estructural de sonidos respiratorios multicanal en pacientes con EPOC para identificar patrones espaciales en la distribución de eventos adventicios, integrando representaciones espectrales, modelado geométrico e interpretabilidad basada en la contribución de las regiones torácicas.

ESPECÍFICOS

- Construir un conjunto de datos multicanal estructurado, integrando metadatos clínicos y características de adquisición para habilitar análisis por canal y por segmento en pacientes con EPOC.
- Extraer representaciones espectrales comparables entre canales, generando un espacio de características que relacione eventos adventicios con configuraciones acústicas multicanal.
- Representar configuraciones espaciales torácicas mediante técnicas geométricas y métodos no supervisados, identificando patrones recurrentes en la distribución regional de sibilancias y crepitantes.
- Desarrollar modelos supervisados e interpretables para la detección de eventos adventicios, evaluando explícitamente la contribución relativa de cada región torácica mediante análisis de importancia y ablation por canal.
- Integrar los hallazgos no supervisados y supervisados para proponer representaciones espaciales explicativas que revelen regularidades en la distribución de eventos, incluso en regiones no etiquetadas.

MARCO CONCEPTUAL

ANALISIS Y CLASIFICACION DE SONIDOS RESPIRATORIOS

El análisis de sonidos respiratorios mediante técnicas de inteligencia artificial ha permitido abordar una amplia variedad de problemas de clasificación, que van desde la identificación de enfermedades específicas hasta la detección de sonidos anómalos y la segmentación en ambientes ruidosos. Entre las aplicaciones más comunes, destaca la clasificación de enfermedades respiratorias como asma, enfermedad pulmonar obstructiva crónica (EPOC), bronquitis, neumonía y COVID-19. Por ejemplo, Chen et al. [1] lograron una precisión del 98.79% al clasificar sonidos normales, estertores y sibilancias utilizando redes ResNets optimizadas, mientras que Fraiwan et al. [2] alcanzaron una precisión del 99.62% al clasificar sonidos normales, asma, neumonía, bronquiectasias y EPOC mediante una combinación de redes neuronales convolucionales (CNN) y memoria a largo plazo bidireccional (biLSTM). En el contexto de la pandemia, el análisis de sonidos respiratorios relacionados con el COVID-19 ha sido un tema central, con varios estudios alcanzando precisiones superiores al 99% gracias a la combinación de técnicas de preprocesamiento y modelos avanzados como ResNets y CNN [7][3].

Además de la clasificación de enfermedades, el análisis también se ha enfocado en la detección de sonidos respiratorios anómalos como estertores, sibilancias, roncus y estridor, que son indicativos de condiciones específicas como fibrosis pulmonar, bronquitis crónica o asma [8][4]. Otros enfoques incluyen la clasificación de sonidos según el tipo de tos (seca o húmeda) para identificar enfermedades subyacentes y la segmentación conjunta de eventos acústicos antes de su clasificación. Por ejemplo, Jung et al. [6] propusieron un modelo basado en redes convolucionales profundas (DS-CNN) para la clasificación precisa de fases respiratorias, integrando características como STFT y MFCC para mejorar la detección. A su vez, la clasificación en ambientes ruidosos ha representado un desafío abordado mediante técnicas de preprocesamiento avanzadas, como la normalización y la extracción de características robustas, que han permitido resultados consistentes incluso en entornos no controlados [1]. Estas contribuciones reflejan el amplio espectro de problemas que el análisis de sonidos respiratorios busca resolver, consolidándose como una herramienta clave para el diagnóstico y monitoreo en salud respiratoria.

La evolución de los modelos aplicados en la clasificación de sonidos respiratorios ha seguido un desarrollo paralelo al avance de las técnicas de inteligencia artificial y aprendizaje automático. Desde enfoques tradicionales basados en modelos estadísticos hasta arquitecturas avanzadas de aprendizaje profundo, se han utilizado diversas aproximaciones para abordar problemas específicos en el análisis de sonidos respiratorios. Inicialmente, los modelos basados en estadística y aprendizaje supervisado dominaron este campo. Entre los primeros estudios destacados, los modelos de máquinas de soporte vectorial (SVM) y k-nearest

neighbors (KNN) fueron ampliamente utilizados debido a su robustez y simplicidad. Por ejemplo, investigaciones en 2018 aplicaron SVM para clasificar sonidos de sibilancias con una precisión del 75.21% [14]. A medida que la complejidad de los problemas aumentaba, también se introdujeron métodos basados en bosque aleatorio (RF) y regresión logística (LR), que ofrecían mayor flexibilidad para analizar características de señales acústicas [7].

En la última década, los modelos de redes neuronales artificiales (ANN) comenzaron a ganar protagonismo, integrando un mayor nivel de automatización en la extracción y clasificación de características. En 2019, estudios como los de Meng et al. [15] demostraron que las ANN eran capaces de superar a los modelos basados en SVM y KNN en tareas de clasificación de sonidos respiratorios, logrando precisiones significativamente mejores gracias a su capacidad para manejar grandes volúmenes de datos y características complejas. En paralelo, los estudios exploraron el uso de redes convolucionales (CNN), destacando su habilidad para identificar patrones en espectrogramas y señales acústicas. Para 2020, investigaciones aplicaron arquitecturas de CNN combinadas con ResNets optimizados para mejorar la clasificación en problemas multicategoría, alcanzando precisiones de hasta el 98.79% [16].

Más recientemente, los avances en aprendizaje profundo han dado lugar a la implementación de arquitecturas más complejas y específicas para tareas de clasificación y segmentación. En 2021, se introdujeron arquitecturas híbridas que combinaban CNN con redes de memoria a largo plazo (LSTM) y memoria a largo plazo bidireccional (biLSTM), logrando una precisión del 99.62% en la identificación de enfermedades respiratorias como asma, EPOC y neumonía [3]. En el mismo año, Hsu et al. [5] propusieron un marco de trabajo comparativo que integraba redes recurrentes con variantes como GRU y CNN-BiLSTM, mostrando mejoras significativas en tareas de detección de fases respiratorias y sonidos adventicios.

A partir de 2022, los esfuerzos se han centrado en explorar modelos más avanzados, como los transformadores y las redes de convolución separable en profundidad (DS-CNN). Jung et al. [6] aplicaron estas arquitecturas en conjunto con características avanzadas como STFT y MFCC para clasificar sonidos respiratorios, alcanzando resultados robustos incluso en condiciones de ruido. Paralelamente, los modelos basados en ResNets optimizados y técnicas de extracción de características avanzadas, como la transformada S, han mostrado una capacidad notable para mejorar la detección de estertores y sibilancias [1]. Este progreso en la implementación de modelos refleja un esfuerzo continuo por integrar técnicas más precisas y adaptativas para el análisis de sonidos respiratorios en contextos clínicos y de investigación.

Los desafíos futuros en la clasificación de sonidos respiratorios incluyen mejorar la interpretabilidad de los modelos de inteligencia artificial, garantizar la calidad y diversidad de los datos utilizados, y desarrollar algoritmos más robustos frente a ruidos y variabilidad en las grabaciones. Además, la integración de tecnologías en dispositivos portátiles y sistemas de telemedicina enfrenta limitaciones en eficiencia computacional y estándares de interoperabilidad. También es crucial establecer metodologías estandarizadas para la adquisición y anotación de datos, facilitando la comparación entre estudios y la validación de modelos. Por último, la implementación en contextos clínicos requiere superar barreras regulatorias y éticas, asegurando cumplimiento normativo y privacidad de los datos, lo que demanda una perspectiva multidisciplinaria para consolidar estas tecnologías como herramientas confiables en la práctica médica

AUSCULTACIÓN Y SONIDOS RESPIRATORIOS

0. Terminología acústica

Para facilitar la comprensión de los sonidos respiratorios y su análisis digital, se presentan a continuación algunos conceptos básicos de acústica y procesamiento de señales.

- **Frecuencia:** La frecuencia es el número de ciclos por segundo de una señal periódica y se mide en hertz (Hz). Determina el tono percibido: frecuencias bajas se asocian a tonos graves y frecuencias altas a tonos agudos.
- **Amplitud:** La amplitud representa la magnitud de la variación de la señal en el tiempo. Está relacionada con la intensidad o energía del sonido.

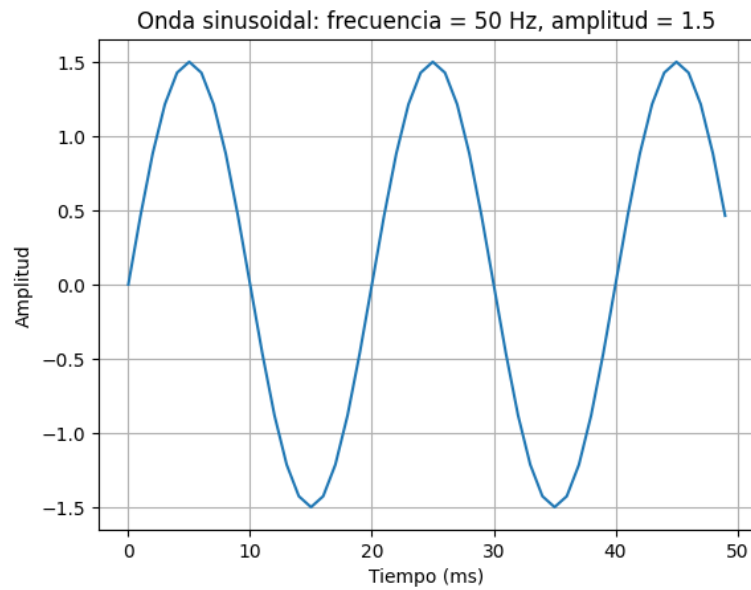


Figura 1: Grafica de onda seno para mostrar amplitud y frecuencia

- Espectro de frecuencia: El espectro describe cómo se distribuye la energía de una señal entre distintas frecuencias. Se obtiene mediante transformadas como la de Fourier y permite analizar el contenido frecuencial de una señal.

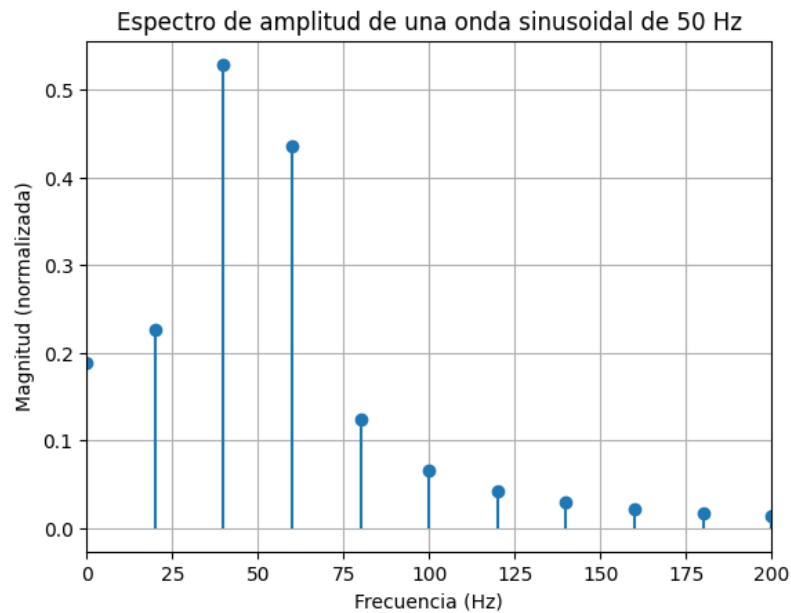


Figura 2: Espectro de frecuencia

- Banda ancha y componentes tonales; Una señal de banda ancha presenta energía distribuida en un rango amplio de frecuencias. Una señal tonal concentra su energía en una o pocas frecuencias bien definidas.
- Duración y estructura temporal: La duración corresponde al intervalo de tiempo en el que ocurre un evento sonoro. La estructura temporal describe si una señal es continua, intermitente o está formada por pulsos breves.
- Timbre: Es una cualidad subjetiva que permite distinguir entre 2 fuentes sonoras de distinta naturaleza. En términos técnicos depende de la combinación de componentes frecuenciales y de su evolución temporal.
- Atenuación: La atenuación es la disminución de la amplitud de una señal al propagarse en un medio, ya sea por absorción, dispersión o pérdidas mecánicas.
- Ruido y relación señal–ruido: El ruido corresponde a componentes no deseados que se superponen a la señal de interés. La relación señal–ruido (SNR) cuantifica cuánta energía de la señal útil hay en comparación con la energía de ruido

1. ¿Qué es la auscultación y cómo funciona este proceso?

La auscultación es una técnica diagnóstica no invasiva que permite evaluar el estado de los órganos internos, especialmente el sistema respiratorio, mediante la escucha de sonidos. Introducida por René Laennec en 1816 con la invención del estetoscopio, esta técnica se centra en identificar patrones acústicos relacionados con anomalías fisiológicas. Su importancia clínica radica en su capacidad para detectar condiciones como neumonía, asma y enfermedad pulmonar obstructiva crónica (EPOC), entre otras [24].

El proceso implica colocar el estetoscopio en diferentes regiones del tórax para escuchar y analizar sonidos respiratorios, que se generan por el flujo turbulento de aire a través de las vías respiratorias. Estos sonidos reflejan características fisiopatológicas como obstrucciones, inflamaciones o secreciones, y se clasifican en normales (tráqueales, vesiculares) y anormales o adventicios (sibilancias, cruídos, estridor) [21].

2. Descripción de los sonidos respiratorios y la superficie torácica

Los sonidos respiratorios son señales acústicas generadas por el flujo de aire a través de las vías respiratorias y por la interacción de dicho flujo con distintas estructuras anatómicas, como la tráquea, los bronquios y el parénquima pulmonar. Desde el punto de vista acústico, se componen de señales de baja amplitud, de banda ancha y carácter cuasi-aleatorio, fuertemente moduladas por la anatomía del tórax y la mecánica respiratoria.

2.1 Sonidos respiratorios normales

Los sonidos respiratorios normales se clasifican habitualmente en:

- a. Sonidos vesiculares: Se auscultan sobre las zonas periféricas del pulmón. Se perciben como suaves, de baja frecuencia (aproximadamente 100–300 Hz) y sin componentes tonales definidos. Reflejan principalmente la transmisión filtrada del flujo turbulento originado en las vías aéreas más grandes hacia las regiones periféricas.
- b. Sonidos bronquiales: Se registran sobre las vías respiratorias mayores o sobre áreas de consolidación pulmonar. Son más intensos, de mayor frecuencia (hasta 600–700 Hz) y con una fase espiratoria relativamente prolongada.
- c. Sonidos traqueales: Se auscultan directamente sobre la tráquea (posición Tc). Presentan alta energía, espectro amplio y timbre áspero, asociados a un flujo altamente turbulento en las vías centrales.

En conjunto, estos sonidos reflejan el flujo normal del aire a lo largo del árbol bronquial y su transmisión a través del tejido pulmonar y la caja torácica.

2.2 Sonidos respiratorios adventicios

Los sonidos respiratorios adventicios son señales adicionales que no deberían estar presentes en condiciones fisiológicas normales. Constituyen marcadores acústicos de alteraciones estructurales o funcionales del aparato respiratorio.

- a. Sibilancias (wheezes): Son sonidos de carácter musical, continuos, con componentes tonales bien definidos. Presentan frecuencias dominantes típicamente en el rango de 400–1000 Hz. Se producen por el estrechamiento de las vías aéreas (broncoespasmo, inflamación, moco), lo que induce vibración de las paredes bronquiales cuando el flujo se vuelve inestable en conductos parcialmente colapsados. Clínicamente se asocian a asma, EPOC y otras formas de obstrucción bronquial.
- b. Crepitantes (crackles): Son sonidos discontinuos, transitorios, de corta duración (del orden de 5–20 ms). Cubren un rango de frecuencias que puede variar aproximadamente entre 100 y 2000 Hz. Suelen clasificarse en crepitantes finos y gruesos. Los finos se asocian frecuentemente a fibrosis pulmonar o edema, mientras que los gruesos se relacionan con secreciones inflamación bronquial.
- c. Estridor: Es un sonido intenso, de tono alto, predominantemente inspiratorio. Se origina en las vías aéreas altas (laringe o tráquea proximal) y está relacionado con obstrucción crítica del calibre de la vía aérea superior.

Otros sonidos, como roncus o frotos pleurales, también pueden aparecer, aunque su representación y frecuencia varían según la patología y el contexto clínico.

La superficie torácica influye directamente en la modulación y transmisión de todos estos sonidos. Variaciones en el grosor de la pared, la masa muscular, la cantidad de tejido adiposo y el estado del parénquima pulmonar pueden alterar la amplitud, el contenido espectral y el timbre de los sonidos, haciendo que ciertas regiones sean más adecuadas para detectar anomalías específicas.

3. Superficie torácica y modulación de los sonidos respiratorios

La superficie torácica actúa como un filtro acústico. La señal que se ausculta en la piel no es el sonido “original” generado en las vías aéreas, sino una versión atenuada y filtrada por:

El grosor de la pared torácica (músculo, grasa, tejido óseo).

- La simetría y morfología del hemitórax.
- El volumen pulmonar y el nivel de insuflación.
- La presencia de consolidaciones, colapso alveolar, líquido o fibrosis en el parénquima.

Por esta razón, la auscultación se realiza de forma sistemática en regiones específicas del tórax:

- Regiones anteriores: supraclaviculares, infraclaviculares, mamarias y submamarias.
- Regiones laterales: áreas axilares (anterior, media y posterior).
- Regiones posteriores: zonas escapulares y subescapulares.

Esta sistematización permite comparar regiones homólogas entre ambos hemitórax y detectar asimetrías en la distribución de los sonidos respiratorios, lo que resulta clave para la identificación de diversas patologías.

4. Posiciones estandarizadas para el registro digital: Tc, Al, Ar, LI, Lr, Pl, Pr

En el contexto del análisis digital de sonidos respiratorios, el grupo CORSA (Computerized Respiratory Sound Analysis), descrito por Sovijärvi et al. [25], propuso un conjunto de posiciones estandarizadas de registro para facilitar la comparabilidad entre estudios y asegurar una cobertura anatómica adecuada. Estas posiciones incluyen:

- Tc (tráquea): región suprasternal o cervical anterior sobre la tráquea.
- Al (anterior izquierdo) y Ar (anterior derecho): campos pulmonares anteriores, típicamente en áreas infraclaviculares o mamarias.
- LI (lateral izquierdo) y Lr (lateral derecho): zonas axilares, representativas de los campos pulmonares medios.

- PI (posterior izquierdo) y Pr (posterior derecho): campos pulmonares posteriores, usualmente a nivel escapular o subescapular.

La elección de estas posiciones obedece a varios criterios:

- a. Cobertura anatómica representativa: Las posiciones anteriores (Al, Ar), laterales (LI, Lr) y posteriores (PI, Pr), junto con la tráquea (Tc), permiten muestrear diferentes regiones del pulmón y de la vía aérea central. De este modo, se cubren campos superiores, medios e inferiores, así como la vía aérea proximal.
- b. Sensibilidad diferencial a distintas patologías: Las regiones posteriores (PI, Pr) son especialmente útiles para detectar crepitantes asociados a patologías que afectan las bases pulmonares, como fibrosis o edema. Las regiones anteriores pueden ser más sensibles a sonidos bronquiales u obstrucciones centrales. La posición Tc capta sonidos de alta energía originados en la tráquea, que sirven de referencia para distinguir entre sonidos generados centralmente y aquellos transmitidos desde la periferia.
- c. Razonamiento acústico y simetría espacial: Cada sitio presenta propiedades de transmisión distintas, pero en conjunto forman una disposición aproximadamente simétrica que permite estudiar:
 - Diferencias entre hemitórax derecho e izquierdo
 - Gradientes dorsoventrales (anterior–posterior).
 - Variaciones laterales.
- d. Comparabilidad y estandarización: El uso sistemático de Tc, Al, Ar, LI, Lr, PI y Pr como posiciones de registro facilita que diferentes investigaciones utilicen un mismo esquema de auscultación digital, favoreciendo la comparación de resultados y la creación de bases de datos compartidas. Esta estandarización es fundamental para el desarrollo y validación de algoritmos de análisis computacional de sonidos respiratorios.

5. Funcionamiento de los sonidos respiratorios desde el punto de vista físico

Desde una perspectiva física, los sonidos respiratorios se originan principalmente por:

- a. Flujo turbulento en las vías aéreas: En las vías aéreas grandes (tráquea y bronquios principales), el flujo de aire se vuelve turbulento, generando vibraciones que se traducen en señales acústicas de banda ancha. Estas señales se atenúan y filtran a medida que se transmiten hacia la periferia pulmonar, dando lugar a los sonidos vesiculares y bronquiales.
- b. Interacción con el parénquima pulmonar: Las vibraciones se propagan a través del parénquima y la pared torácica. Cambios en la densidad o estructura del tejido (por ejemplo, consolidaciones, fibrosis o edema)

modifican la forma en que estas vibraciones se transmiten, alterando el timbre y la amplitud de los sonidos auscultados.

- c. Fenómenos específicos asociados a sonidos adventicios: En el caso de los crepitantes, los mecanismos incluyen la apertura abrupta de pequeñas vías aéreas colapsadas o fenómenos de burbujeo en presencia de líquido. En las sibilancias, el estrechamiento de los bronquios genera condiciones de flujo inestable, con vibración de las paredes bronquiales que produce un sonido tonal característico. En el estridor, la obstrucción de las vías altas crea un flujo de alta velocidad a través de un estrechamiento, generando un sonido intenso y de alta frecuencia.

El espectro y la morfología temporal de los sonidos respiratorios dependen, por tanto, de una combinación de factores: calibre de las vías, fase respiratoria, volumen pulmonar, propiedades mecánicas del parénquima y posición de registro sobre la superficie torácica. Esta complejidad justifica el uso de técnicas de análisis digital y modelos de ciencia de datos para caracterizar de forma más objetiva y reproducible los sonidos respiratorios, especialmente los adventicios. [8] [24] [21].

6. Implementación de estetoscopios electrónicos

Los estetoscopios electrónicos han evolucionado como una extensión de los dispositivos convencionales, permitiendo la grabación, análisis y almacenamiento de sonidos respiratorios. Estos dispositivos integran tecnología digital para:

- Mejorar la calidad del sonido mediante la amplificación y filtrado de ruidos.
- Facilitar el análisis computacional mediante la extracción de características acústicas.
- Integrar algoritmos de inteligencia artificial para identificar patrones acústicos y correlacionarlos con patologías específicas [8].

Además, los estetoscopios electrónicos han expandido el acceso a la atención médica mediante soluciones portátiles conectadas a redes inalámbricas, permitiendo consultas remotas y telemedicina.

La modulación de estos sonidos está influenciada por factores como la resistencia de las vías respiratorias, el volumen pulmonar y la ubicación anatómica, lo que destaca la importancia de una evaluación sistemática de la superficie torácica.

DESCRIPTORES DE AUDIO PARA LA CARACTERIZACIÓN ACÚSTICA

1. Transformada de Fourier (TF)

La Transformada de Fourier (TF) es una técnica matemática que descompone una señal en el dominio del tiempo en sus componentes de frecuencia. Esto se logra representando la señal como una combinación de senoidales con diferentes frecuencias, amplitudes y fases. Es ampliamente utilizada para analizar señales estacionarias y proporciona una representación espectral completa (Lerch, 2012).

Matemáticamente, la TF está definida como:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt$$

donde $X(f)$ es el espectro de la señal $x(t)$, f representa la frecuencia, y $e^{-j2\pi ft}$ descompone la señal en sus componentes sinusoidales.

La TF es eficaz para estudiar señales estacionarias donde las características de frecuencia no cambian con el tiempo. Sin embargo, presenta limitaciones significativas al analizar señales no estacionarias, como los sonidos respiratorios, donde las características espectrales cambian dinámicamente en función del tiempo.

2. Limitaciones de la TF y la Transformada de Fourier de Tiempo Corto (STFT)

Para señales no estacionarias, la TF tradicional no captura adecuadamente las variaciones temporales. Por esta razón, se utiliza la Transformada de Fourier de Tiempo Corto (STFT). Este enfoque divide la señal en segmentos temporales (ventanas) y aplica la TF a cada segmento. Esto permite un análisis tiempo-frecuencia, proporcionando una representación dinámica que combina información espectral y temporal [12].

La STFT está definida como:

$$X(t, f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j2\pi f\tau} d\tau$$

donde $w(\tau - t)$ es una ventana deslizante que define el intervalo temporal analizado.

La elección del tamaño de la ventana es crucial: ventanas más pequeñas ofrecen mejor resolución temporal pero peor resolución espectral, y viceversa. La STFT es especialmente útil para analizar señales dinámicas como las respiratorias, que combinan componentes estacionarias y transitorias.

3. Descriptores Acústicos

Los descriptores acústicos son métricas extraídas del espectro para caracterizar propiedades específicas de una señal. A continuación, se describen los más relevantes para el análisis de sonidos respiratorios:

- Spectral Slope: Mide la inclinación de la envolvente espectral, indicando cómo se distribuye la energía entre bajas y altas frecuencias. Es útil para diferenciar entre señales de timbre oscuro y brillante.

$$Slope = \frac{\sum ((f_k - \bar{f})(X_k - \bar{X}))}{\sum ((f_k - \bar{f})^2)}$$

- Spectral Crest Factor: Relación entre el pico máximo y la energía promedio del espectro, evaluando la prominencia de ciertas frecuencias.

$$Crest = \frac{\max(X_k)}{\left(\sum \frac{X_k}{N}\right)}$$

- Spectral Centroid: Es el "centro de gravedad" del espectro, indicando la frecuencia promedio ponderada por la amplitud.

$$Centroid = \frac{\sum (f_k \cdot X_k)}{\sum (X_k)}$$

- Spectral Spread: Mide la extensión del espectro alrededor del centroide espectral, representando su dispersión.

$$Spread = \sqrt{\frac{\sum ((f_k - Centroid)^2 \cdot X_k)}{\sum (X_k)}}$$

- Spectral Skewness: Describe la asimetría de la distribución espectral. Un espectro sesgado hacia frecuencias altas tendrá un skewness positivo.

$$Skewness = \frac{\left[\frac{\sum ((f_k - Centroid)^3 \cdot X_k)}{\sum (X_k)} \right]}{Spread^3}$$

- Spectral Kurtosis: Evalúa la concentración de energía alrededor de las frecuencias centrales. Valores altos indican un espectro con picos muy pronunciados.

$$Kurtosis = \frac{\left[\frac{\sum ((f_k - Centroid)^4 \cdot X_k)}{\sum (X_k)} \right]}{Spread^4}$$

- Spectral Roll-Off: Define la frecuencia por debajo de la cual se encuentra un porcentaje determinado de la energía espectral acumulada (por ejemplo, el 85%).

$$\Sigma(X_k) \text{ para } f_k \leq f_r = p \cdot \Sigma(X_k)$$

- Spectral Flatness: Relación entre la media geométrica y la media aritmética del espectro. Valores bajos indican un espectro tonal; valores altos indican ruido.

$$Flatness = \frac{(\prod X_k)^{\frac{1}{N}}}{\left(\Sigma \frac{X_k}{N}\right)}$$

- Spectral Decrease: Representa la disminución de la amplitud del espectro hacia frecuencias más altas.

$$Decrease = \frac{\left[\Sigma \left(\frac{X_k - X_0}{k} \right) \right]}{\Sigma \left(\frac{1}{k} \right)}$$

- Spectral Flux: Mide el cambio entre cuadros consecutivos del espectro, capturando la dinámica del sonido.

$$Flux(n) = \Sigma (X_{n[k]} - X_{\{n-1\}[k]})^2$$

- Spectral Entropy: Mide la uniformidad de la distribución espectral, útil para evaluar la aleatoriedad o estructura de la señal.

$$Entropy = - \Sigma P[k] \log(P[k]), \quad P[k] = \frac{X_k}{\Sigma X_j}$$

Estos descriptores ofrecen una representación compacta de las características acústicas y son esenciales para tareas de clasificación y análisis en señales respiratorias [12].

Por otra parte, otro conjunto de descriptores muy usado en audio son los coeficientes cepstrales de frecuencia de mel,

Los coeficientes cepstrales en las frecuencias de Mel (MFCC) son descriptores que representan la envolvente espectral de una señal en una escala de frecuencia cercana a la percepción humana. Para calcularlos, la señal de tiempo continuo se segmenta en tramas cortas y a cada trama se le aplica una ventana para reducir efectos de borde. Sobre cada trama se calcula la transformada rápida de Fourier

(FFT) y se obtiene el espectro de potencia $|X(k)|^2$. Después, ese espectro se pasa por un banco de filtros triangulares espaciados en la escala Mel: las posiciones de los filtros se obtienen transformando las frecuencias lineales f (en Hz) a la escala Mel mediante la relación:

$$mel(f) = 2595 \cdot \log_{10}(1 + f / 700)$$

El resultado de aplicar cada filtro m es una energía espectral filtrada $E(m)$, que se puede escribir como:

$$E(m) = \sum_k (|X(k)|^2 \cdot H_m(k))$$

donde $H_m(k)$ es la respuesta del m -ésimo filtro Mel en el bin espectral k . A continuación, se aplica un logaritmo a estas energías, $L(m) = \log(E(m))$, modelando la respuesta aproximadamente logarítmica del oído. Finalmente, se aplica la transformada discreta del coseno (DCT) a la secuencia $L(m)$ para obtener los MFCC, que son una especie de “cepstrum sobre escala Mel”. El coeficiente cepstral de orden c se expresa como:

$$MFCC(c) = \sum_{m=1}^M L(m) \cdot \cos\left(\pi \cdot c \cdot \frac{(m - 0.5)}{M}\right)$$

donde M es el número de filtros Mel y $c = 0, 1, \dots, C-1$ es el índice del coeficiente cepstral. Estos coeficientes resumen la forma de la envolvente espectral de cada trama y, por su estabilidad y capacidad discriminante, se usan de forma habitual en análisis de voz, sonidos respiratorios y otras señales biomédicas. [12]

REDUCCIÓN DE DIMENSIONALIDAD CON ESCALAMIENTO MULTIDIMENSIONAL CLÁSICO (MDS) Y ANÁLISIS DE PROCRUSTES

1. Escalamiento Multidimensional Clásico (MDS)

El Escalamiento Multidimensional Clásico (MDS) es una técnica estadística que permite representar datos de alta dimensionalidad en un espacio de menor dimensión, preservando las distancias o disimilitudes entre los puntos. Su objetivo es facilitar la visualización y comprensión de la estructura subyacente de los datos.

Procedimiento del MDS Clásico:

- Construcción de la Matriz de Distancias: Se calcula una matriz que contiene las distancias o disimilitudes entre cada par de objetos.
- Centrado Doble de la Matriz: Se aplica una transformación para centrar la matriz de distancias, ajustando la escala de los datos.

- Descomposición Espectral: Se realiza una descomposición en valores y vectores propios de la matriz centrada para identificar las direcciones principales de variación.
- Selección de Dimensiones: Se eligen las dimensiones correspondientes a los mayores valores propios, y los datos se proyectan en este subespacio de menor dimensión.

Esta metodología fue desarrollada por Warren S. Torgerson en 1958, quien introdujo el modelo métrico clásico de MDS. Posteriormente, Shepard [23] y Kruskal [10] [11] ampliaron el enfoque al proponer modelos no métricos que permiten trabajar con datos en escalas ordinales.

La manera de medir la calidad de estos procesos es mediante la métrica de Stress normalizado.

$$\text{Stress normalizado} = \sqrt{\frac{\sum (d_{original} - d_{reducida})^2}{\sum (d_{original}^2)}}$$

Donde $d_{original}$ y $d_{reducida}$ son respectivamente, la distancia de cada punto en el espacio original y la distancia en el nuevo espacio de dimensión reducida.

Kruskal en [10] definió umbrales para la métrica del Stress normalizado como:

- Excelente: menor a 0.05
- Bueno: entre 0.05 y 0.1
- Aceptable: entre 0.1 y 0.2
- Dudoso: entre 0.2 y 0.5
- Netamente malo: mayor a 0.5

2. Análisis de Procrustes

El Análisis de Procrustes es una técnica utilizada para comparar y alinear configuraciones de puntos en el espacio, eliminando diferencias debidas a traslación, rotación y escala. Su objetivo es superponer dos o más configuraciones de manera óptima para evaluar su similitud en forma y tamaño.

Procedimiento del Análisis de Procrustes:

- Traslación: Se trasladan las configuraciones para que sus centroides coincidan en el origen del sistema de coordenadas.
- Escalado: Se ajustan las configuraciones para que tengan una escala común, generalmente normalizando la suma de las distancias al cuadrado al origen.

- Rotación: Se aplica una rotación óptima que minimiza la distancia entre las configuraciones, comúnmente utilizando descomposición en valores singulares para encontrar la matriz de rotación adecuada.

El término "Procrustes" proviene de la mitología griega, donde Procrustes era un bandido que ajustaba a la fuerza a sus víctimas a la longitud de su cama, simbolizando la idea de forzar una correspondencia entre formas. El Análisis de Procrustes ha sido ampliamente utilizado en morfometría y análisis de formas desde las décadas de 1980 y 1990 [20].

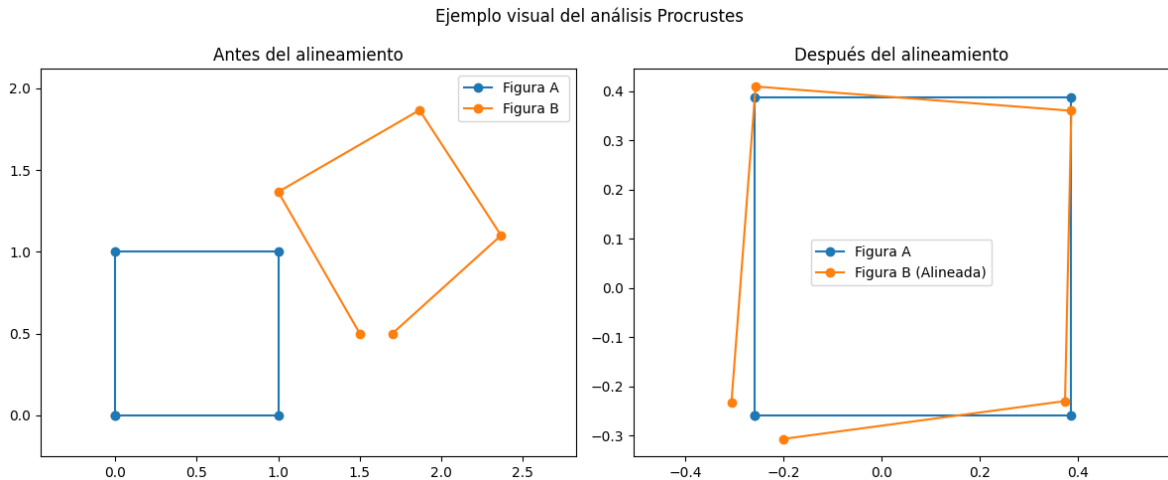


Figura 3: Ejemplo de alineamiento Procrustes

Del análisis Procrustes surge el concepto de distancia Procrustes. Es una medida que evalúa el grado de similitud entre dos configuraciones geométricas (por ejemplo, puntos en el espacio), luego de aplicar transformaciones geométricas permitidas: traslación, escalado y rotación. Es ampliamente utilizada para comparar formas en campos como morfometría y análisis espacial. Dadas dos matrices "X" y "Y", cada una de tamaño n se aplica primero un centrado, escalado y rotación óptima. La distancia Procrustes se define como:

$$D^2 = 1 - \frac{\text{traza}(B^T A A^T B)}{\|A\|_F^2 \cdot \|B\|_F^2}$$

Donde:

- A y B son versiones escaladas y centradas de X y Y ,
- $\text{traza}(\cdot)$ es la suma de la diagonal de la matriz,
- $\|\cdot\|_F$ es la norma de Frobenius

Esta distancia permite medir la calidad del alineamiento y la consistencia espacial.

MODELOS NO SUPERVISADOS

Los modelos no supervisados son algoritmos de aprendizaje automático que operan sin etiquetas de clase. Su objetivo es descubrir estructuras, patrones o agrupaciones presentes en los datos únicamente a partir de sus relaciones internas. Este tipo de métodos se utiliza cuando no existe información previa sobre categorías o cuando se busca explorar la organización natural de los datos. Entre sus aplicaciones se incluyen la detección de patrones, la reducción de dimensionalidad y el agrupamiento (cluster analysis) [40].

Algoritmo k-means

K-means es uno de los algoritmos de agrupamiento más utilizados. Su objetivo es dividir un conjunto de datos en k grupos, de modo que los puntos dentro de cada grupo sean lo más similares posible entre sí y lo más diferentes posible respecto a los puntos de otros grupos. El algoritmo funciona iterativamente, alternando dos pasos:

1. Asignación de cada punto al centroide más cercano.
2. Re-cálculo de los centroides como la media de los puntos asignados a cada grupo.

El proceso continúa hasta que los centroides dejan de cambiar significativamente. K-means es eficiente computacionalmente y es ampliamente usado en análisis exploratorio, segmentación y descubrimiento de patrones.

MODELOS SUPERVISADOS

Los modelos supervisados son algoritmos de aprendizaje automático que requieren un conjunto de datos anotado compuesto por pares entrada–salida. El objetivo es aprender una función que permita predecir etiquetas para nuevas observaciones a partir de ejemplos previos [30]. En aplicaciones biomédicas, estos modelos se emplean comúnmente para la clasificación de eventos fisiológicos y el análisis de señales, incluyendo sonidos respiratorios.

Dentro de los modelos supervisados más utilizados se encuentran los árboles de decisión, máquinas de soporte vectorial, modelos de ensamble y redes neuronales artificiales [28]. Estos métodos han mostrado eficacia en contextos donde las relaciones entre variables pueden ser no lineales, ruidosas o altamente complejas.

Random Forest

Random Forest es un modelo de ensamble basado en la combinación de múltiples árboles de decisión entrenados sobre subconjuntos de los datos mediante muestreo bootstrap. En cada nodo de cada árbol, el algoritmo considera únicamente un subconjunto aleatorio de las características disponibles, lo cual introduce variabilidad y reduce el sobreajuste [29].

La predicción final se obtiene a partir de la votación mayoritaria en tareas de clasificación o del promedio de predicciones en tareas de regresión. Este enfoque permite capturar interacciones no lineales y mejorar la estabilidad del modelo en comparación con árboles individuales.

Importancia de características en Random Forest

Una ventaja del modelo Random Forest es que permite cuantificar la contribución relativa de cada característica al desempeño del modelo. Dos métricas ampliamente utilizadas son:

1. **Importancia por reducción de impureza (Gini importance)**
Se calcula a partir de la disminución de impureza en los nodos donde una característica induce particiones. Esta importancia se promedia entre todos los árboles del bosque [29].
2. **Importancia basada en permutaciones**
Propone medir la caída en el rendimiento del modelo cuando los valores de una característica se permutan aleatoriamente. Este método, introducido originalmente por Breiman [29], está menos sesgado y refleja mejor la relevancia real de cada variable [34].

Estas medidas son especialmente relevantes en aplicaciones biomédicas que requieren comprender la contribución de diferentes señales fisiológicas o regiones anatómicas en la predicción de eventos clínicos.

Métricas de evaluación

La evaluación del rendimiento de un modelo supervisado requiere métricas que cuantifiquen su capacidad de clasificar correctamente las observaciones. En problemas biomédicos, donde las clases suelen estar desbalanceadas, es necesario utilizar métricas más informativas que la exactitud tradicional [33].

- **Precisión:** Mide la proporción de predicciones positivas que son correctas. Se utiliza cuando es importante reducir falsos positivos. Definición clásica en clasificación binaria descrita por Powers (2011).
- **Sensibilidad o recall:** Evalúa la capacidad del modelo para detectar correctamente los casos positivos. Es especialmente relevante cuando los falsos negativos tienen un alto costo, como en aplicaciones clínicas. Definición formal según Powers [32].

- **Medida F1 (F1-score):** Es la media armónica entre precisión y recall. Es apropiada en escenarios con clases desequilibradas y cuando se busca un equilibrio entre ambos errores [33].

DISEÑO METODOLÓGICO

Para la presente investigación se utiliza el marco metodológico CRISP-DM (Cross-Industry Standard Process for Data Mining), el cual proporciona una estructura clara y estandarizada para el desarrollo de proyectos de minería de datos y análisis avanzado. Este enfoque es adecuado para cumplir con los objetivos de este estudio, que busca desarrollar un modelo integrado para la caracterización e interpretación de sonidos respiratorios. A continuación, se detallan las primeras etapas de esta metodología, alineadas con los objetivos planteados.

1. ENTENDIMIENTO DEL PROBLEMA

El primer paso del marco CRISP-DM se enfoca en comprender el problema desde un punto de vista clínico y científico, estableciendo las metas generales y específicas del proyecto.

- **Contexto del Problema:** La auscultación pulmonar es una herramienta clínica esencial para el diagnóstico de enfermedades respiratorias, pero el análisis tradicional presenta limitaciones importantes relacionadas con la subjetividad del examinador y la falta de reproducibilidad. El análisis computacional de sonidos respiratorios ha permitido avances notables en clasificación automática, pero persisten retos asociados a la variabilidad introducida por la posición de grabación y a la limitada interpretabilidad clínica de los modelos actuales. Esta investigación propone abordar dicho vacío mediante el uso de registros multicanal, lo que permite explorar la dimensión espacial de los sonidos respiratorios y comprender cómo se distribuyen los eventos adventicios en diferentes regiones del tórax.
- **Propósito del Estudio:** Diseñar una estrategia de análisis para registros multicanal de pacientes con EPOC que integre: características espectrales,

representaciones geométricas alineadas (Procrustes y variantes), métodos de agrupamiento para la detección de patrones espaciales, y modelos supervisados complementados con técnicas de interpretabilidad. El propósito es no solo identificar la ocurrencia de eventos adventicios (sibilancias y crepitantes), sino también generar representaciones más robustas e informativas que permitan detectar patrones regionales y comprender en qué zonas del tórax se manifiestan con mayor intensidad los fenómenos respiratorios patológicos.

- Resultados Esperados:
 - Un pipeline analítico capaz de extraer y representar características relevantes de los sonidos respiratorios.
 - La identificación de patrones espaciales multicanal asociados con la distribución de sibilancias y crepitantes.
 - Modelos supervisados con capacidad de detección de eventos acústicos y técnicas de interpretabilidad que revelen la contribución relativa de cada región torácica.
 - Una propuesta metodológica robusta que reduzca el impacto de la variabilidad posicional y ofrezca representaciones más cercanas a la práctica clínica.

2. ENTENDIMIENTO DE LOS DATOS

Fuente de datos

Los datos fueron recopilados de la base de datos ICBHI 2017 Respiratory Sound que contiene 920 registros correspondientes a 126 pacientes. Las señales están anotadas según los ciclos respiratorios y la ocurrencia de sibilancias y crepitantes. Hay 6898 ciclos, de los cuales 1864 están etiquetados como crepitantes, 886 como sibilancias, 506 contienen ambos eventos y 3642 no contienen eventos. Además, se dispone de información sobre los trastornos pulmonares de los pacientes, la posición del pecho, el dispositivo de grabación, el modo de adquisición y datos demográficos. La tabla resume las características de la base de datos.

Tabla 1: Descripción de los datos

Dispositivos	WelchAllyn Stethoscope (127),	Meditron AKG C417L (646),	Master 3M	Elite	Electronic
--------------	----------------------------------	------------------------------	--------------	-------	------------

	Littmann Classic II SE Stethoscope (87), 3M Littmann 3200 Electronic Stethoscope (60)
Metodo de grabación	Single channel (188), Multi channel (732)
Patología	Asthma (1), Bronchiectasis (7), Bronchiolitis (6), COPD (64), Healthy (26), LRTI (2), Pneumonia (6), URTI (14)
Genero	Female (46), Male (79), NA (1)
Rangos de edad	0–20 years (50), 20–60 years (11), 60–100 years (64), NA (1)
Rangos de masa corporal	BMI < 18.5 (36), 18.5 < BMI < 25 (28), 25 < BMI < 30 (38), BMI > 30 (15), NA (9)
Pacientes	126
Grabaciones	920

Organización de archivos

Los registros de audio se encuentran en formato **.wav** y siguen una convención de nomenclatura estructurada:

`<IDpaciente>_<IDgrabación>_<parteTórax>_<modoAdquisición>_<dispositivo>.wav`

Por ejemplo:

- 101_1b1_Pr_sc_Meditron.wav
 - 101: identificador del paciente (rango 101–226).
 - 1b1: identificador de la grabación.
 - Pr: parte del tórax (Ar = anterior derecho, Al = anterior izquierdo, Lr = lateral derecho, Ll = lateral izquierdo, Pr = posterior derecho, Pl = posterior izquierdo, Tc = tráquea).
 - sc/mc: método de adquisición (sc = single channel, mc = multichannel).
 - Meditron: dispositivo de grabación utilizado.

Cada grabación tiene asociado un archivo de anotaciones `.txt` que contiene los eventos adventicios detectados, indicando el tiempo de inicio y fin dentro del registro. Por ejemplo:

```
event  start  end
wheeze 0.75 3.48
```

Archivos de metadatos

Adicionalmente, se dispone de un archivo tabular de texto que relaciona el identificador del paciente con su condición clínica, por ejemplo:

```
101    URTI
102    Healthy
103    Asthma
104    COPD
...
```

Este archivo permite asociar a cada paciente con un diagnóstico médico, lo que facilita la construcción de subconjuntos de análisis

3. PREPARACIÓN DE LOS DATOS

Los datos brutos se preprocesan mediante las siguientes operaciones:

- **Remuestreo:** Las señales se remuestrean a un valor uniforme de 4 kHz.
- **Segmentación:** Las señales se segmentan en ciclos respiratorios. Cada ciclo se recorta para eliminar muestras de baja amplitud en los bordes. Los ciclos resultantes con una duración inferior a 75 ms se descartan.
- **Normalización:** Sea $\tilde{x}_i$ el ciclo respiratorio original, se obtiene un nuevo x_i aplicando normalización estándar a $\tilde{x}_i$.

Ventana Deslizante

El análisis de las señales respiratorias se realizó mediante la aplicación de una ventana deslizante de longitud N_w , con un porcentaje de solapamiento definido como α_w . Esta configuración permitió dividir la señal en segmentos que pueden considerarse aproximadamente estacionarios, una condición necesaria para una representación consistente de las dinámicas de la señal. El tamaño de la ventana (N_w) se ajusta para que cada segmento capture condiciones estacionarias, mientras

que el porcentaje de solapamiento (α_w) controla la resolución temporal, asegurando una cobertura suficiente de la señal sin perder información crítica sobre sus transiciones.

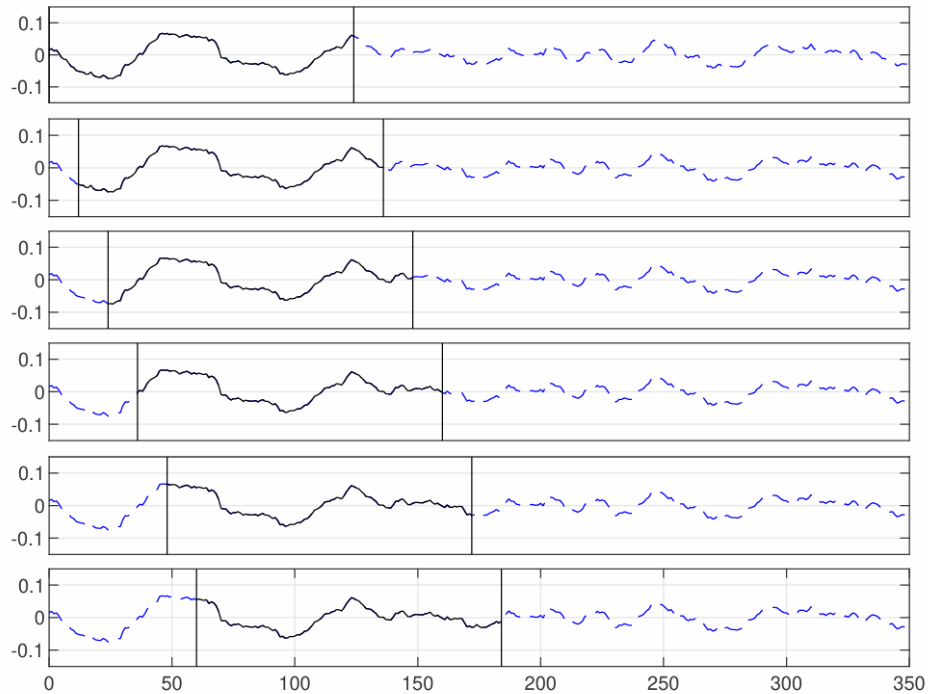


Figura 4: Ejemplo de segmentación de una señal de audio

Cada uno de estos segmentos cortos de tiempo se usará como una observación.

Extracción de descriptores espectrales

En este estudio, las características se extraen mediante dos tipos de análisis: un análisis temporal y un análisis espectral.

- **Análisis espectral:** Se extrae un conjunto de 11 características del espectro de las señales, a saber: centroide, dispersión, asimetría (skewness), curtosis, factor de cresta, roll-off, planitud espectral, taper, flujo (flux), pendiente (slope) y entropía.
- **Análisis MFCC:** Se extraen 13 características usando los coeficientes cepstrales de frecuencia de mel.

El análisis espectral se realiza en 3 bandas de frecuencia definidas como baja (60 Hz a 300 Hz), media (300 Hz a 800 Hz) y alta (800 Hz a 1800 Hz). En la figura 5 se

observa una característica extraída en 3 bandas de frecuencia distintas. Teniendo en cuenta las representaciones espectrales y temporales, se obtienen un total de 33 características, que se referencian mediante el índice p. Las Tablas 2 y 3 enumeran las características extraídas y su índice p.

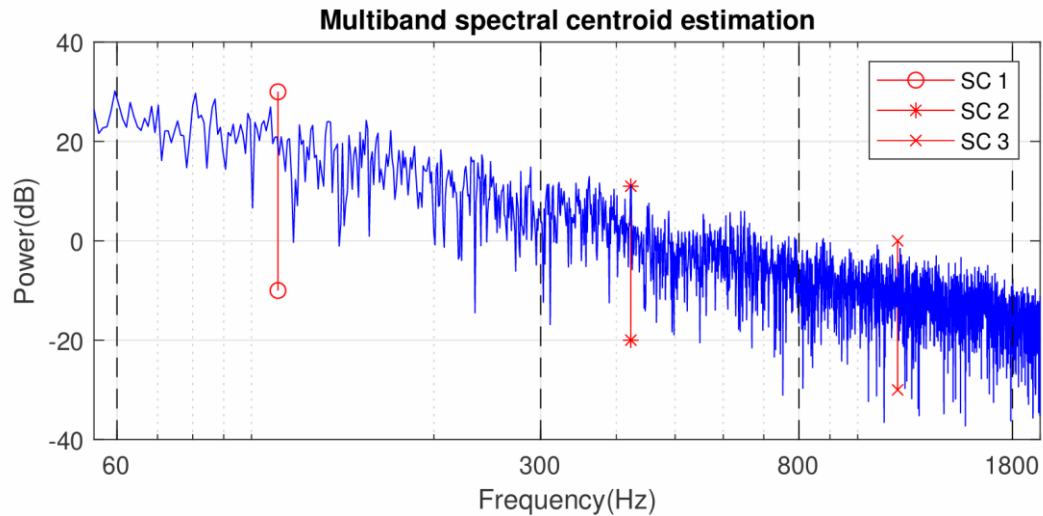


Figura 5: Extracción de descriptor acústico en 3 bandas espectrales

Tabla 2: Descriptores para cada banda de frecuencia

Descriptor	Índice
Spectral slope	1, 2, 3
Spectral Crest Factor	4, 5, 6
Spectral skewness	7, 8, 9
Spectral kurtosis	10, 11, 12
Spectral spread	13, 14, 15
Spectral roll off	16, 17, 18
Spectral flatness	19, 20, 21
Spectral decrease	22, 23, 24
Spectral flux	25, 26, 27
Spectral centroid	28, 29, 30
Spectral entropy	31, 32, 33

Tabla 3: Descriptores MFCC

Descriptor	Índice
MFCC 1-13	34-46

Construcción de datasets

Se seleccionaron únicamente las grabaciones **multicanal registradas con el dispositivo AKG C417L**, que incluyen simultáneamente los siete puntos torácicos: anterior derecho (Ar), anterior izquierdo (Al), lateral derecho (Lr), lateral izquierdo (Li), posterior derecho (Pr), posterior izquierdo (Pl) y tráquea (Tc).

Dada la naturaleza multicanal de los datos, se construyeron dos estructuras analíticas diferenciadas:

1. Dataset para análisis geométrico (Procrustes y clustering):

- Para cada segmento, se generó un vector de **46 características por canal**, lo que produce una configuración de **7 vectores**.

2. Dataset para aprendizaje supervisado (árboles de decisión):

- La columna de “parte del tórax” se **pivotó** para disponer las 46 características espectrales de cada canal como columnas, resultando en un vector de **322 características (46×7)** por segmento.

Consideraciones sobre calidad de datos

- Se incluyeron únicamente segmentos con los siete canales completos en el análisis geométrico, para garantizar configuraciones consistentes.

En la figura 6 se muestra a groso modo como es el flujo del proceso de preparación de los datos.

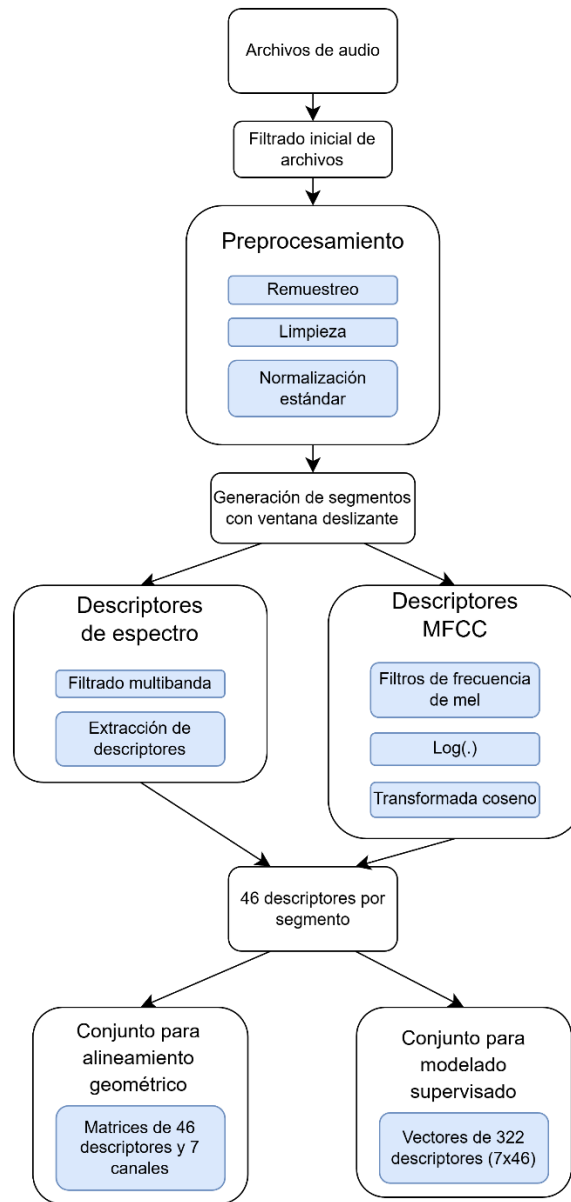


Figura 6: Proceso seguido para la extracción y la transformación de datos

4. MODELADO DE DATOS

Alineamiento geométrico

Cada segmento de audio multicanal fue representado como una configuración geométrica compuesta por siete vectores (uno por canal torácico), cada uno de ellos con 46 descriptores espectrales. Para estudiar relaciones espaciales entre estos vectores se implementó una estrategia de reducción y alineamiento geométrico

seguida de un agrupamiento no supervisado, con el objetivo de detectar patrones recurrentes en la distribución de eventos acústicos en el tórax.

Escalamiento multidimensional

Para facilitar la visualización y reducir la dimensionalidad, se aplicó un escalamiento multidimensional (MDS) a cada configuración multicanal. El MDS se realizó sobre los vectores de 46 dimensiones de cada canal, proyectándolos a un espacio de 2 o 3 dimensiones, según el análisis. Esta transformación preserva, en la medida de lo posible, las relaciones de disimilitud entre los canales, generando así **configuraciones de 7 vectores con 2 características** para cada segmento multicanal.

Se escogió este algoritmo en concreto por el hecho de basarse en preservar las distancias de los vectores de alta dimensión. Esto hace que, en el nuevo espacio de baja dimensión, se mantienen las relaciones de disimilitud entre vectores, algoritmos como tsne o umap pueden romper esta relación.

Para manejar esta aproximación, se tiene que elegir una noción de distancia, en este caso se probaron 2, una distancia euclídeana y una logarítmica absoluta. La distancia euclídeana se escogió por ser la distancia en la que se basa en escalamiento clásico [10]. Por otra parte, la distancia logarítmica absoluta se escogió porque lidia mejor con altas dimensiones y se ha probado su buen funcionamiento para reducir la métrica del Stress en este tipo de problemas [37].

Análisis de Procrustes

Para representar configuraciones espaciales de manera coherente entre segmentos multicanal, se implementó un procedimiento de alineamiento geométrico basado en la distancia de Procrustes. Este procedimiento permite comparar figuras o conjuntos de puntos en 2D eliminando los efectos de rotación, traslación y escala.

En el flujo de trabajo esto se realizó sobre las dimensiones reducidas por el escalamiento dimensional, se hace así debido a que trabajar Procrustes con muchas dimensiones En marcos de datos muchas dimensiones el problema se vuelve “ill-posed”, lo que significa que hay muchas transformaciones equivalentes (rotaciones/reflexiones) que ajustan casi igual. No hay una solución única con interpretación confiable [43].

Para Alinear un conjunto completo de puntos se necesita una configuración de referencia, en este proyecto se planteó un algoritmo de para extraer un **centroide o prototipo de un conjunto de configuraciones** basado en Generalized Procrustes Analysis (GPA) o el cálculo de la Media de Fréchet en el espacio de formas [41] [42]. El algoritmo consiste en los siguientes pasos:

- a. **Inicialización:** Se define como centroide inicial la mediana punto a punto de todas las configuraciones sin alinear. Esta elección (en lugar del promedio) ofrece mayor robustez frente a valores atípicos.
- b. **Iteración de alineamiento:** En cada iteración, todas las configuraciones se alinean a la referencia actual (el centroide) utilizando el análisis Procrustes clásico. Esto genera un conjunto de configuraciones alineadas y una lista de distancias de Procrustes asociadas a cada alineación.
- c. **Actualización del centroide:** Se calcula una nueva referencia como la **mediana punto a punto** de las configuraciones alineadas obtenidas en el paso anterior. A su vez, se computa la mediana de las distancias de alineación como medida de convergencia.
- d. **Criterio de convergencia:** El proceso iterativo continúa hasta que el cambio en la mediana de las distancias sea menor a una tolerancia predefinida (tol) o comience a aumentar (indicando sobreajuste). En tal caso, se detiene el proceso y se conserva el último centroide estable.
- e. **Salida:** El algoritmo retorna tres elementos clave:
 - Las configuraciones alineadas finales.
 - El centroide o prototipo (figura promedio) resultante.
 - Las distancias Procrustes individuales de cada configuración al centroide.

Este proceso permite representar cada segmento multicanal como una figura estandarizada y comparable, facilitando su posterior agrupamiento o su análisis como entidad topológica. Además, las distancias Procrustes resultantes sirven como medidas cuantitativas del grado de desviación de cada segmento con respecto a la estructura promedio del grupo.

En la figura 7 se muestra el diagrama de flujo que explica el algoritmo de alineamiento ala referencia centroide.

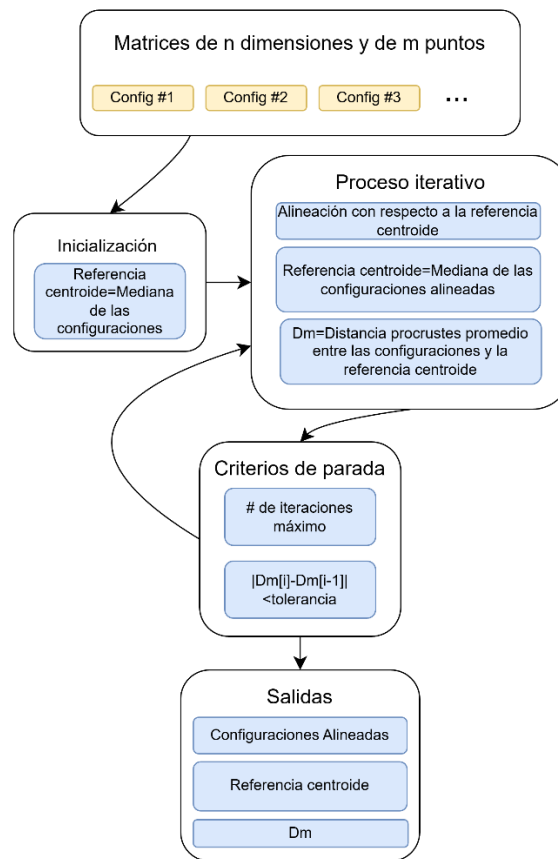


Figura 7: Proceso de algoritmo de alineamiento a referencia centroide

Agrupamiento Procrustes

Para identificar configuraciones espaciales recurrentes en los sonidos respiratorios multicanal, se implementó un esquema de agrupamiento personalizado que combina la **alineación Procrustes** con una estrategia iterativa tipo **K-Means**, utilizando como métrica la **distancia de Procrustes** y como noción de centroide una **media geométrica iterativa** obtenida mediante el procedimiento de extracción de centroide descrito anteriormente.

Cada configuración (segmento multicanal) se representa como una figura de siete puntos en el espacio proyectado por MDS. Estas figuras se alinean entre sí respecto a una referencia común, permitiendo comparar configuraciones estructurales independientemente de su posición, rotación o escala.

a. Inicialización de centroides

Para evitar resultados sensibles a la inicialización, se utilizó una variante adaptada del método **K-Means++**, en la cual los centroides se seleccionan

progresivamente de acuerdo con su distancia mínima (Procrustes) a los centroides ya elegidos. Esta técnica favorece la diversidad estructural entre los centroides iniciales, mejorando la estabilidad y calidad del agrupamiento.

b. Iteraciones de agrupamiento

Durante el proceso iterativo:

- Cada configuración es asignada al centroide más cercano según la distancia de Procrustes.
- Se recalculan los centroides de cada grupo aplicando el algoritmo de alineamiento centroide, que encuentra la figura media de las configuraciones del cluster.
- El procedimiento se repite hasta alcanzar un número máximo de iteraciones.

Esta estrategia permite identificar patrones de similitud espacial entre las configuraciones multicanal, y constituye una representación topológica más robusta frente a la variabilidad posicional y anatómica de los registros. El proceso se encuentra descrito en la figura 8

Para la implementación de este método se creó una case llamada “proclustering” con métodos:

- **fit:** Calcula k referencias centroides
- **predict:** Recibe un conjunto de configuraciones y les asigna una etiqueta dependiendo de la referencia centroide más cercana
- **Inertia:** Calcula la inercia del resultado para una muestra de configuraciones.

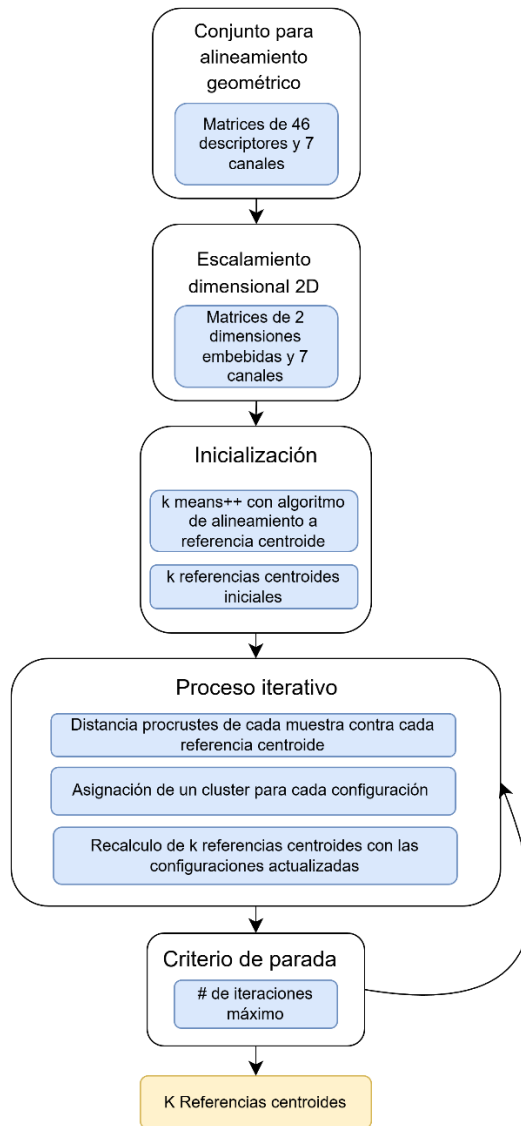


Figura 8: Proceso de estimación de k referencias centroides

Modelado Supervisado

Para entrenar modelos supervisados, se construyó un conjunto de datos que representa cada segmento multicanal como una única observación de 322 características. Estas características se obtienen al pivotar la variable “parte del tórax” para reorganizar las 46 características espectrales por canal, resultando en un vector de tamaño $46 \times 7 = 322$. Esta representación permite capturar simultáneamente la información espectral en todas las regiones

torácicas medidas (Al, Ar, Ll, Lr, Pl, Pr, Tc), lo que facilita el análisis conjunto de la distribución espacial de los sonidos respiratorios.

La etiqueta de cada segmento corresponde a la ocurrencia de eventos adventicios (sibilancias, crepitantes o ausencia de eventos) y se asigna a todo el segmento si al menos un canal contiene una anotación en ese intervalo. Esto permite evaluar si señales aparentemente limpias en ciertos canales podrían seguir aportando información predictiva.

Selección del tipo de modelo:

Se optó por utilizar **Random Forest (RF)** como modelo base debido a las siguientes razones:

- **Robustez frente al ruido y datos redundantes:** RF es un ensamble de árboles que mitiga el sobreajuste al promediar múltiples árboles entrenados sobre subconjuntos aleatorios de datos y características.
- **Adaptabilidad a datos multivariados:** Su capacidad para manejar datos de alta dimensión sin requerir supuestos lineales lo hace especialmente útil para los vectores multicanal espectrales utilizados aquí.
- **Interpretabilidad:** A diferencia de modelos más complejos como redes neuronales profundas, RF permite extraer fácilmente métricas de importancia de características, lo cual es crucial para los objetivos de interpretabilidad regional y espectral de este estudio.
- **Rendimiento competitivo sin ajuste intensivo:** RF logra buenos resultados en tareas de clasificación sin una necesidad crítica de ajuste de hiperparámetros, lo que favorece la estabilidad del modelo en estudios exploratorios como este.

Aunque podrían haberse usado modelos más simples, como regresión logística o K-vecinos más cercanos, estos tienden a tener dificultades en presencia de relaciones no lineales complejas o de colinealidad entre variables. En cambio, se privilegió un balance entre interpretabilidad y capacidad de modelado.

Experimentos adicionales

Se realizaron análisis de **ablación**, en los cuales se comparó el rendimiento del modelo al excluir canales específicos, como la tráquea, con el fin de evaluar su aporte en la detección de eventos. Los resultados mostraron cambios significativos en las métricas para ciertas clases, como las sibilancias, lo que sugiere que incluso regiones no etiquetadas directamente pueden contener información útil.

5. EVALUACIÓN

La etapa de evaluación en el marco CRISP-DM se enfoca en valorar la calidad de los modelos generados y la utilidad de los hallazgos para los objetivos definidos. En este proyecto, se consideran dos tipos de evaluación: una para los modelos de agrupamiento no supervisados, y otra para los modelos de clasificación supervisada.

Evaluación de modelos no supervisados (Proclustering)

El método de clustering basado en la distancia de Procrustes y alineamiento geométrico fue evaluado mediante métricas adaptadas a la noción de forma y configuración espacial. Dado que el algoritmo utiliza alineamiento Procrustes como su noción de similitud y promedio, no es adecuado utilizar métricas tradicionales como la inercia euclidiana o el coeficiente de silueta clásico.

Se adoptaron las siguientes métricas:

- **Inercia Procrustes:** definida como la suma de las distancias Procrustes entre cada observación y su centroide del clúster correspondiente.
- **Índice de Calinski-Harabasz modificado:** adaptado para utilizar la noción de distancia Procrustes y los centroides obtenidos mediante el alineamiento

Estas métricas permitieron evaluar la compacidad y separación entre clústeres sin asumir distribuciones gaussianas ni estructuras lineales. Adicionalmente, se realizaron análisis post-clustering para cuantificar la densidad de eventos acústicos por clúster, con el fin de identificar patrones espaciales clínicamente significativos.

Evaluación de modelos supervisados

Para los modelos supervisados contruidos con Random Forest, se utilizaron métricas estándar que permiten una evaluación robusta frente a clases desbalanceadas y multiclase:

- **Precision, Recall y F1-score** macro-promediados para las clases wheezes, crackles y none.

Estas métricas se calcularon tanto para el modelo general como para los modelos entrenados por clúster. Este enfoque permitió detectar si ciertos agrupamientos derivados del análisis geométrico presentan mejores condiciones para la clasificación de eventos acústicos.

También se aplicaron experimentos de **ablación**, removiendo información de canales específicos (por ejemplo, tráquea), con el fin de observar el impacto en el

rendimiento del modelo. Esto permitió identificar qué regiones torácicas aportan de forma crítica al proceso de predicción.

Evaluación de interpretabilidad

Se utilizó la **importancia de características** derivada del Random Forest para obtener medidas de interpretabilidad. Se analizaron:

- Las características más relevantes por canal torácico.
- Las características más relevantes por banda de frecuencia.
- La variación de importancia entre clústeres.

Esto permitió establecer no solo qué descriptores espectrales son más informativos, sino también en qué regiones del cuerpo su contribución es más significativa. Además, se analizó el uso de las distancias a los centroides de los clústeres como una nueva representación embebida para evaluación posterior.

RESULTADOS Y DISCUSIÓN

CONSTRUCCIÓN DE CONJUNTOS DE DATOS

Selección y Filtrado Inicial de Registros Multicanal

Para este estudio se seleccionaron únicamente registros multicanal grabados con el dispositivo AKG C417L, ya que permiten capturar señales simultáneamente desde siete regiones torácicas. Tras el filtrado, se obtuvieron 224 archivos de audio correspondientes a 32 grabaciones multicanal, cada una con una duración uniforme de 20 segundos.

Duración de los Eventos Acústicos

Para comprender mejor su dinámica temporal, se construyeron histogramas de duración:

- **Wheezes:** Tienen una duración media de ~496 ms, pero muestran un sesgo con una concentración importante de eventos alrededor de 200 ms.
- **Crackles:** Mucho más breves, con una media de ~26 ms y un máximo registrado de aproximadamente 100 ms.

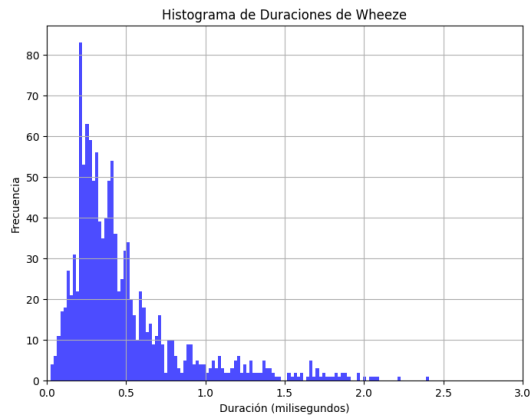


Figura 9: Histogramas de duración de eventos acústicos wheeze

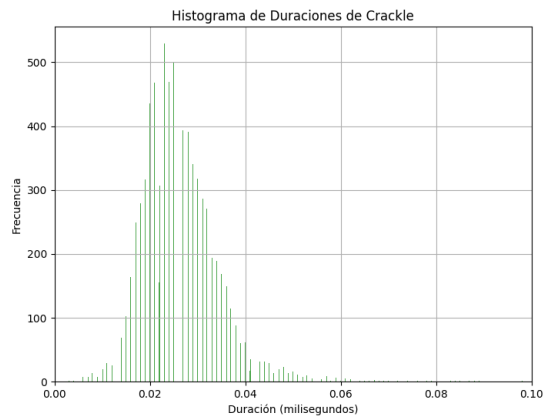


Figura 10: Histograma de duración de eventos acústicos crackle

Segmentación de la Señal: Ventana Deslizante

Con base en esta distribución, se definió una segmentación temporal mediante una ventana deslizante de:

- Tamaño de ventana (N_w): 100 ms
- Solapamiento (ϕ_w): 40%

Esta elección balancea varios factores:

- Permite capturar crackles completos (aun si su duración media es menor).
- Es tolerante al corte de wheezes, dada su naturaleza más estacionaria.
- Mantiene una buena estabilidad de las características espectrales al evitar ventanas demasiado pequeñas

La segmentación de los 224 audios generó un total de 74,368 segmentos individuales. Estos se agruparon para formar 10,624 grabaciones multicanal segmentadas, cada una correspondiente a una ventana de análisis multicanal.

Eventos por Segmento y por Región Torácica

Luego de la segmentación, se identificaron eventos dentro de las ventanas, resultando en:

- 5251 segmentos con wheezes
- 5313 segmentos con crackles

Estos números son con respecto los 74,368 segmentos individuales.

El siguiente resumen muestra la distribución de eventos detectados por región torácica (Nc = crackles, Nw = wheezes):

Tabla 4: Conteo de eventos por región torácica

Posición Torácica	Crackles	Wheezes
Al	620	769
Ar	972	644
LI	1233	1467
Lr	935	595
PI	1282	702
Pr	1011	474
Tc	687	709

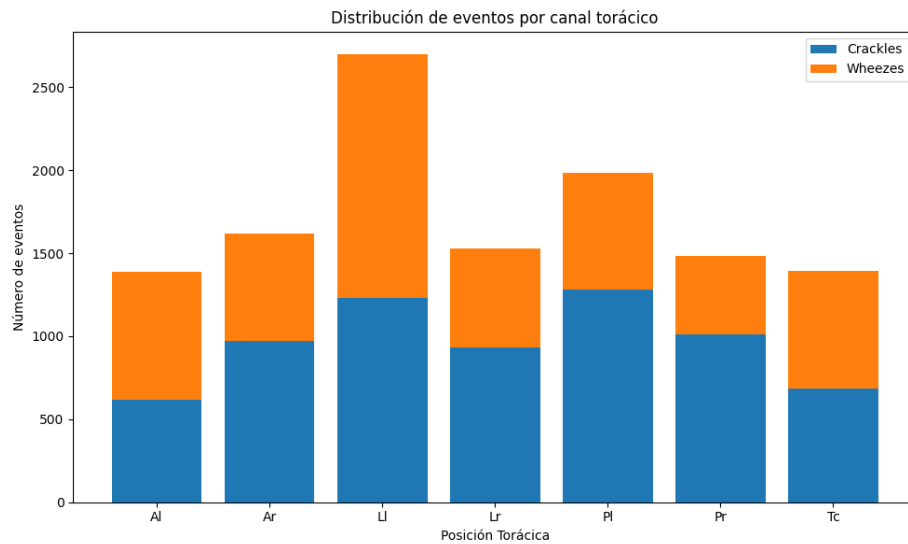


Figura 11: Distribución de eventos por región torácica

En la figura 11, se observa que los eventos crackles tienden a acumularse en zonas posteriores e izquierdas del tórax (especialmente PI y LI), lo cual es consistente con patrones clínicos comunes en enfermedades como la EPOC. Por otro lado, los wheezes presentan un comportamiento más disperso, aunque también destacan en LI, Al, y Tc.

EXTRACCIÓN DE CARACTERÍSTICAS

Tras la segmentación y limpieza de los datos, se extrajeron **46 características por canal**, combinando descriptores espectrales y 13 coeficientes cepstrales (MFCCs). Estas características buscan capturar aspectos relevantes del contenido acústico de los eventos adventicios.

En su espectro, las señales muestran comportamientos acústicos distintos dependiendo de la parte del cuerpo. Por ejemplo, en la figura 12 se muestra como en la parte anterior del tórax se percibe un **crackle**, mientras que en la parte posterior se distinguen 2 muy marcados.

También se proyectaron las observaciones en un espacio PCA en la figura 13, únicamente con fines ilustrativos. En esta proyección no se ve ninguna separación evidente, además de ningún patrón relacionado con la región del tórax. Esto refuerza la necesidad de usar otro enfoque para poder detectar patrones espaciales en las señales multicanal, ya que, si se comparan las características en un espacio euclidiano y lineal, no se logra extraer información relevante.

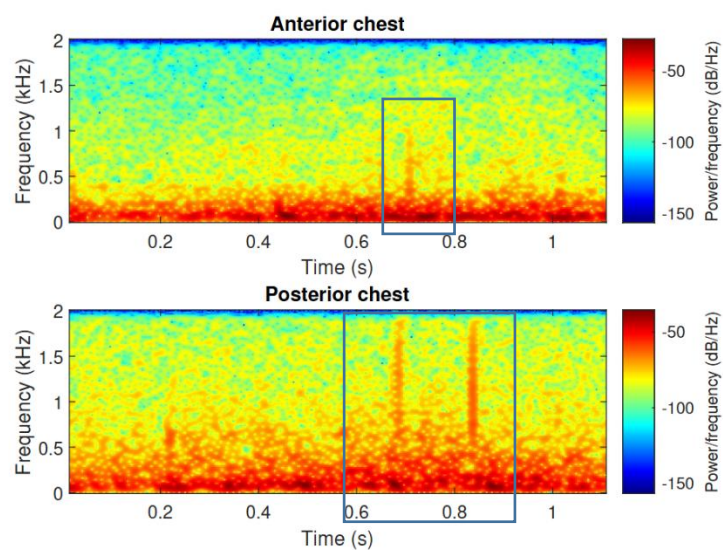


Figura 12: Espectrograma de 2 canales respiratorios en donde ocurren crackles

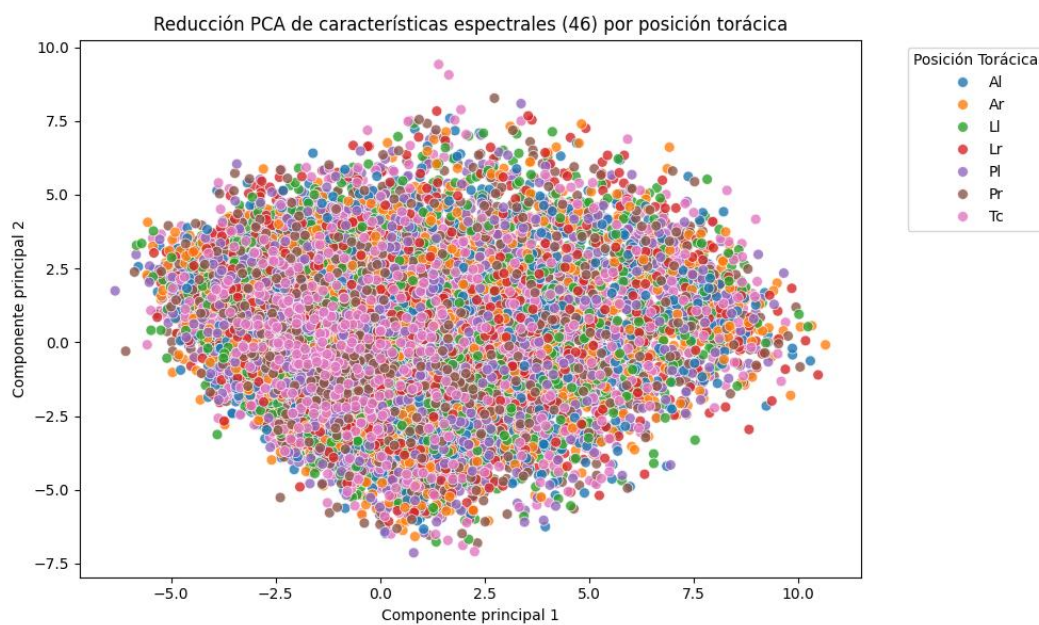


Figura 13: Proyección con PCA de las en un espacio lineal

Finalmente se evaluó la correlación de pearson de las características para descartar variables con una correlación muy alta, el umbral escogido para este caso fue de

0.9 y las variables eliminadas fueron las de menor varianza. En la figura 14 se muestra el mapa de calor de las correlaciones.

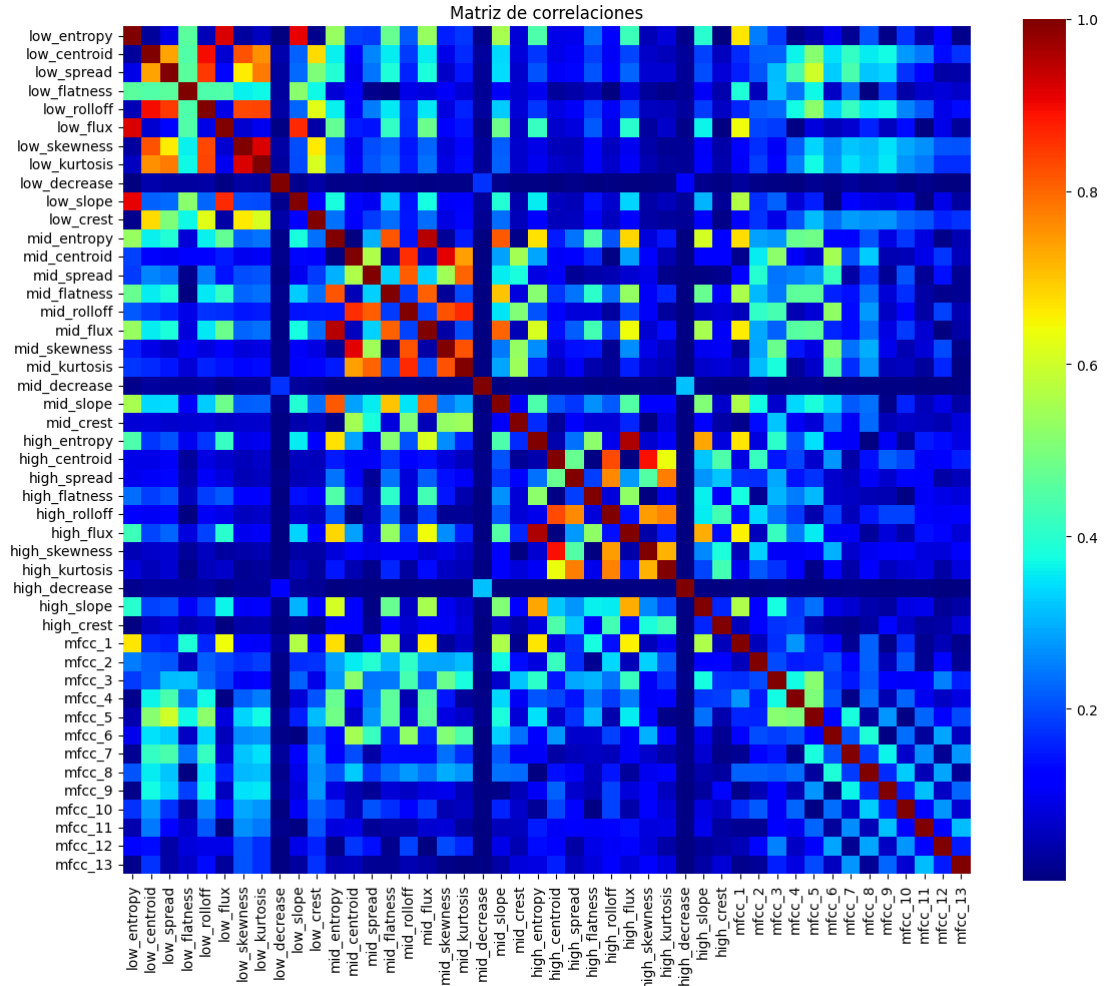


Figura 14: Mapa de calor de correlación de variables

De las 46 características originales, se descartaron 14, 11 del análisis espectral y 3 del análisis cepstral

Variables descartadas: 'mid_rolloff', 'mid_flux', 'mid_skewness', 'mid_kurtosis', 'mid_slope', 'high_flatness', 'high_rolloff', 'high_flux', 'high_skewness', 'high_kurtosis', 'high_slope', 'mfcc_5', 'mfcc_6', 'mfcc_7'

ALINEAMIENTO GEOMETRICO

Proyecciones de segmentos multicanal en un espacio 2D

En la figura 15 se refleja el escalamiento dimensional del conjunto global de datos (74368 segmentos individuales) debido a limitaciones computacionales, se tomó una muestra de 30000 segmentos. Aquí se plasma que hacer el escalamiento de los datos sin distinguir la pertenencia a un mismo segmento multicanal, no permite visualizar patrones útiles, además de tener un Stress alto de **0.3**.

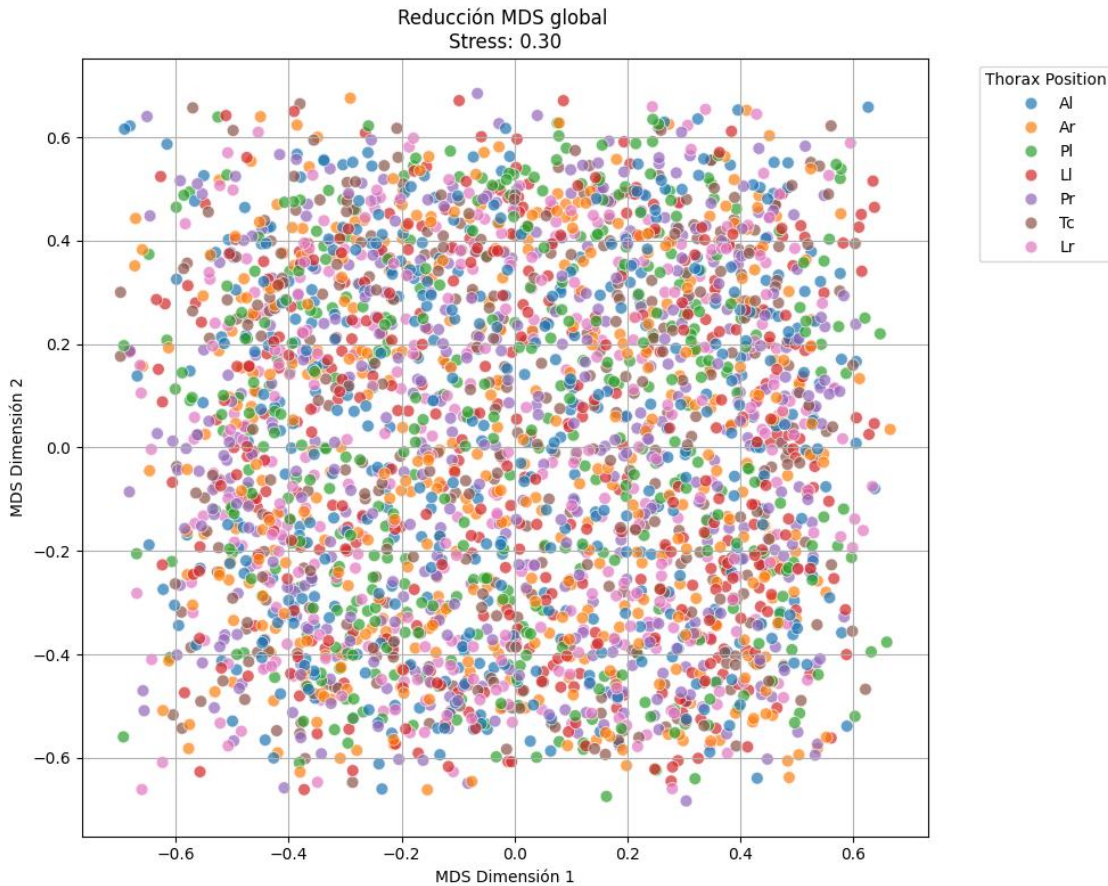


Figura 15: Escalamiento multidimensional sobre el conjunto de 74368 segmentos individuales

Teniendo esto en cuenta, toman las 10624 configuraciones de 7 vectores, cada uno de 32 dimensiones, obtenidos en la etapa de extracción de características. A cada segmento multicanal se le hizo escalamiento con distancia euclidiana y distancia logarítmica absoluta. La noción de distancia escogida para proceder con el experimento es la **logarítmica absoluta** ya que como se ve en la figura 17, hace que la distribución del Stress normalizado se concentre en valores “buenos” y “aceptables” [10]. Por otra parte, en la figura 17, la distribución de Stress para la

distancia euclidiana se centra más en el umbral de valores “dudosos” por lo cual los escalamientos con esta distancia están siendo de peor calidad.

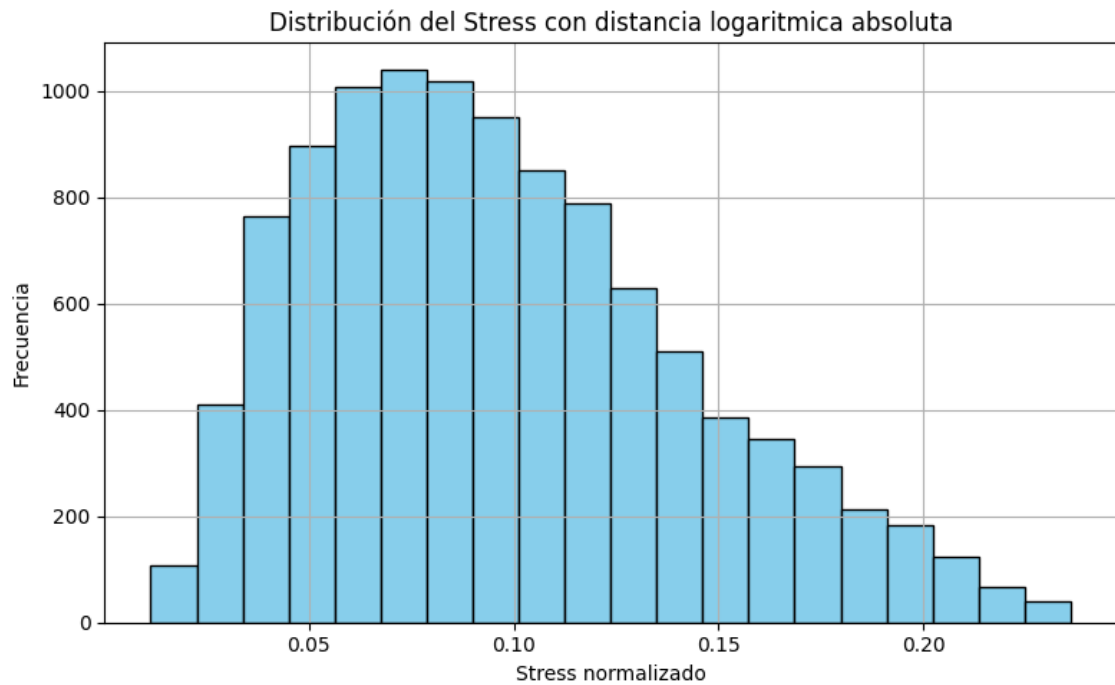


Figura 16: Distribución del Stress normalizado para distancia logarítmica absoluta

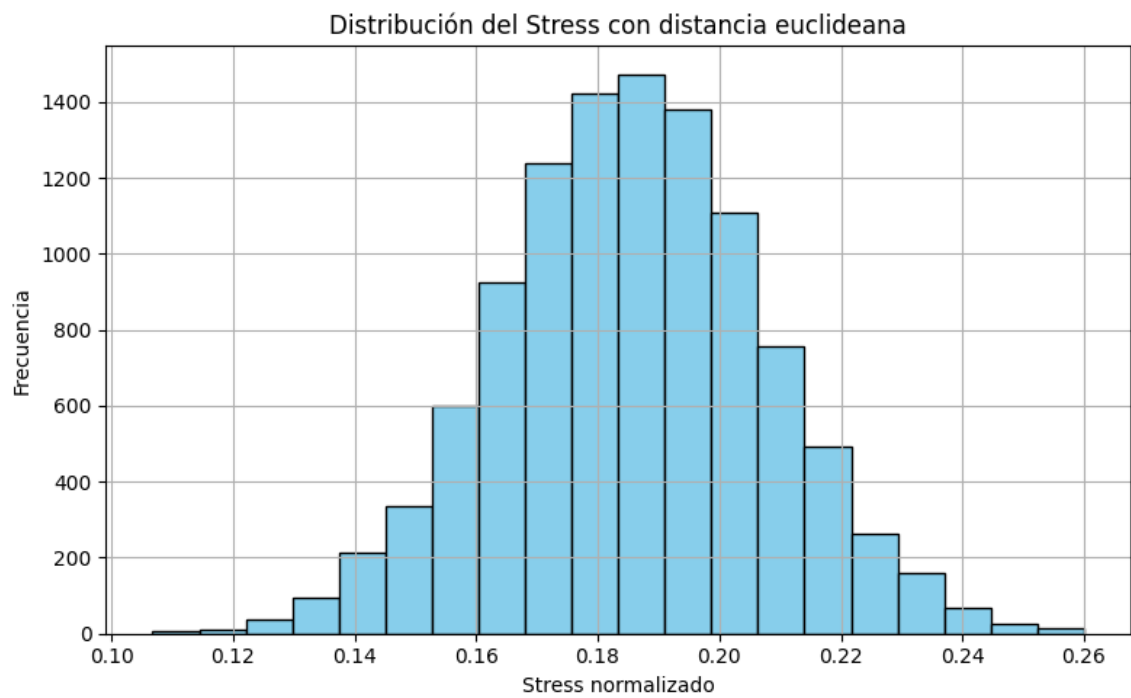


Figura 17: Distribución del Stress normalizado para distancia euclideana

En la figura 18 se ilustra un segmento multicanal proyectado en un espacio bidimensional mediante escalamiento multidimensional clásico donde los puntos representan las posiciones anatómicas: tráquea, regiones laterales, anteriores y posteriores. En este ejemplo en concreto destaca la distancia entre la tráquea y los canales del tórax, además que hay más concentración entre los canales de la izquierda (**PI**, **LI**, **AI**) y más dispersión en los canales derechos (**Pr**, **Lr**, **Ar**).

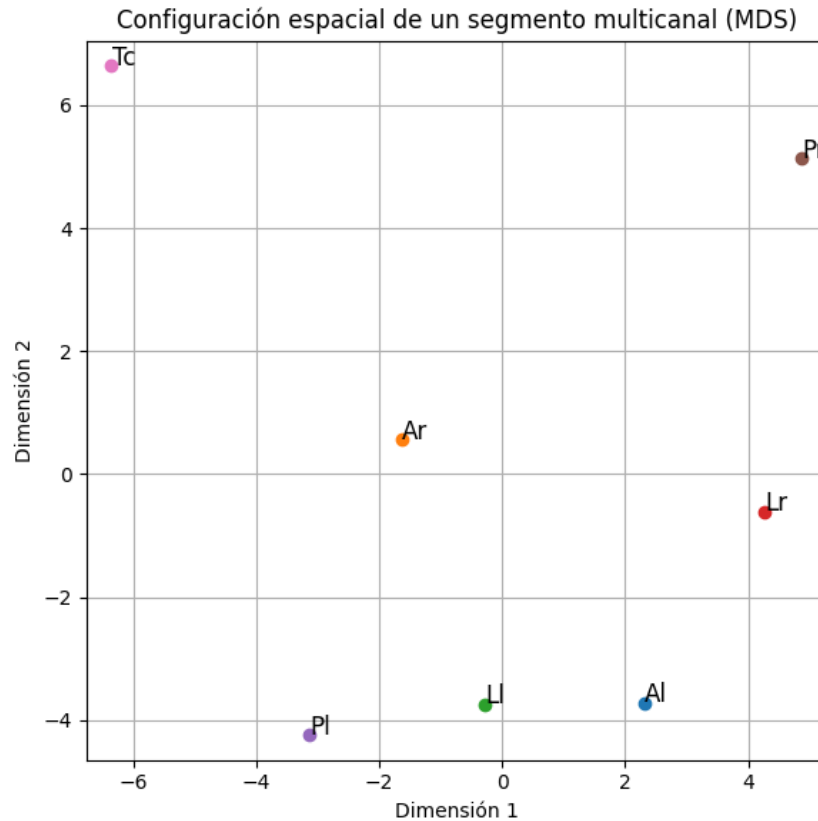


Figura 18: Segmento multicanal procesado por escalamiento dimensional

Alineamiento con Procrustes

Una vez proyectadas las configuraciones multicanal en un espacio bidimensional mediante escalamiento multidimensional clásico (MDS), se aplica un proceso de alineamiento Procrustes. En las figuras 19 y 20 se muestra el resultado de alinear las configuraciones a esa referencia centroide para todo el conjunto de datos. En las dos figuras se contrastan las muestras cercanas a esta referencia promedio y las más distantes, el criterio de usado para esto es la distancia Procrustes.

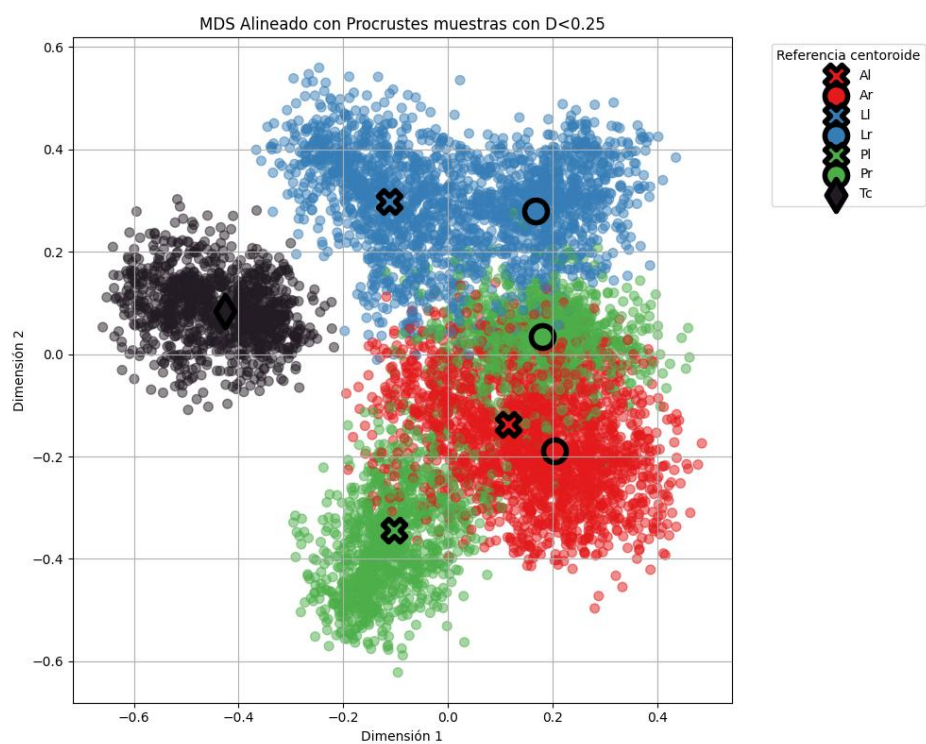


Figura 19: Alineamiento promedio de todo el conjunto de datos, configuraciones con $D < 0.25$

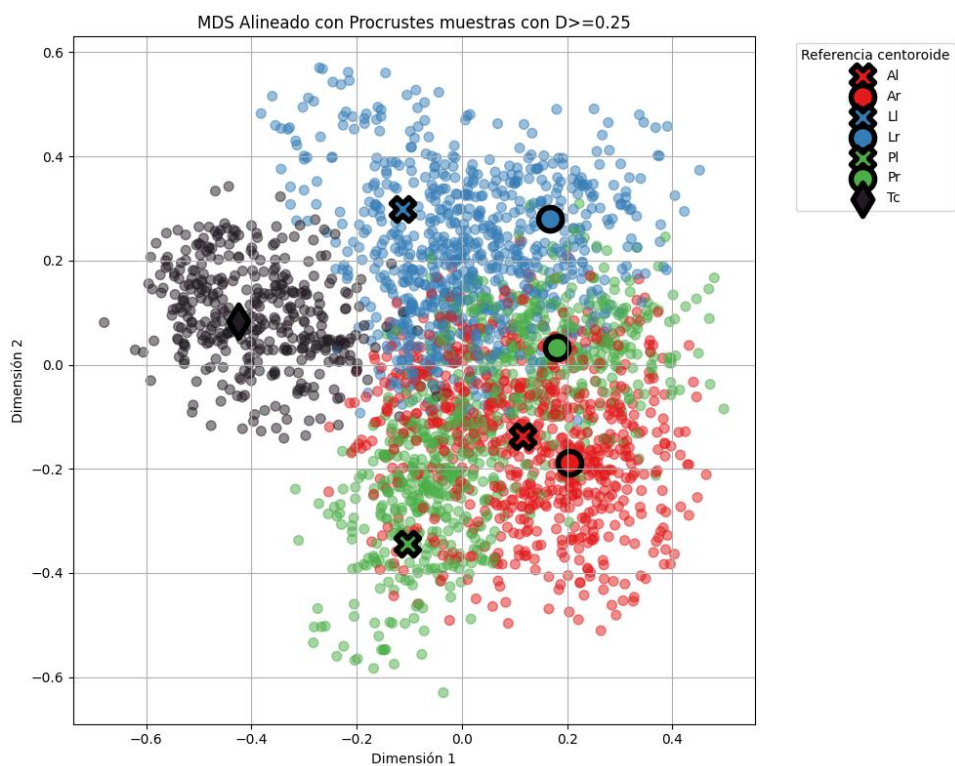


Figura 20: Alineamiento promedio de todo el conjunto de datos, configuraciones con $D \geq 0.25$

Este contraste permite apreciar que, si bien muchas configuraciones se alinean de forma consistente, otras presentan divergencias más marcadas que justifican la posterior aplicación de un proceso de agrupamiento específico.

En las pruebas realizadas de este método se observó que los alineamientos son visiblemente buenos cuando la distancia Procrustes promedio está por debajo de **0.25** [36].

Proclustering

Como primer paso, se explora el comportamiento de la inercia en función del número de clústeres con el objetivo de estimar un valor adecuado de k . En la figura 21 se muestra la correspondiente gráfica del codo, que sugiere un punto de inflexión alrededor de **$k=3$** , pero por motivos experimentales se decide usar **$k=6$** , ya que este mantiene la distancia promedio de cada punto a su clúster por debajo de **0.22** y posteriormente al analizar la distancia promedio de cada clúster, todos están por debajo de **0.25**.

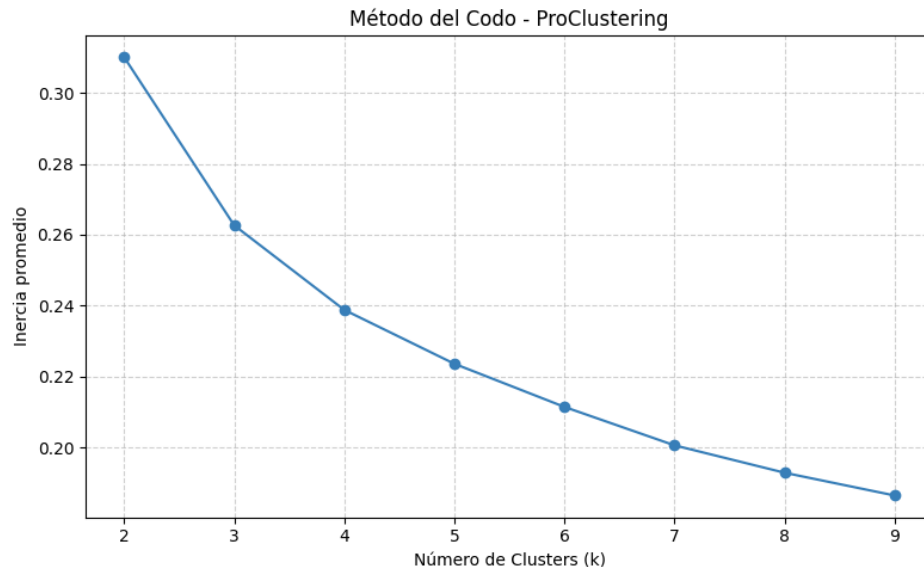


Figura 21: Gráfica del codo para inercia promedio

Representación espacial de los clústeres

Las figuras 22 a 27 muestra los datos de cada clúster, obtenidos tras alinear las configuraciones individuales con sus respectivos centroides. Estas configuraciones reflejan distribuciones geométricas de eventos acústicos en el tórax. Aunque comparten una estructura general coherente, se observan diferencias sutiles en la posición relativa y densidad espacial de las regiones torácicas. Un patrón general es que la información de **Tc** es distante a todos los puntos anatómicos en casi todos los clústeres, esto tiene sentido ya que la tráquea como tal no hace parte del tórax, en ella se perciben sonidos vocales además de los respiratorios.

Respecto a las demás regiones torácicas, en cada clúster se tiene una distribución diferente:

- Clúster 0: Es el único en donde **Tc** se concentra con otras regiones del tórax, destaca como se alejan los puntos **Pr** y **Pl**.
- Clúster 1: **Tc** está alejado y las regiones torácicas se concentran y se mezclan.
- Clúster 2: En este grupo **Tc** también se aleja, las regiones **Lr** y **Li** se concentran
- Clúster 3: Similar al anterior grupo, con la diferencia de que **Pl** se aleja.
- Clúster 4: Similar al clúster 1, con la diferencia de que aquí **Pr** se aleja.
- Clúster 5: Este grupo tiene el mismo comportamiento que la referencia promedio de todos los datos.

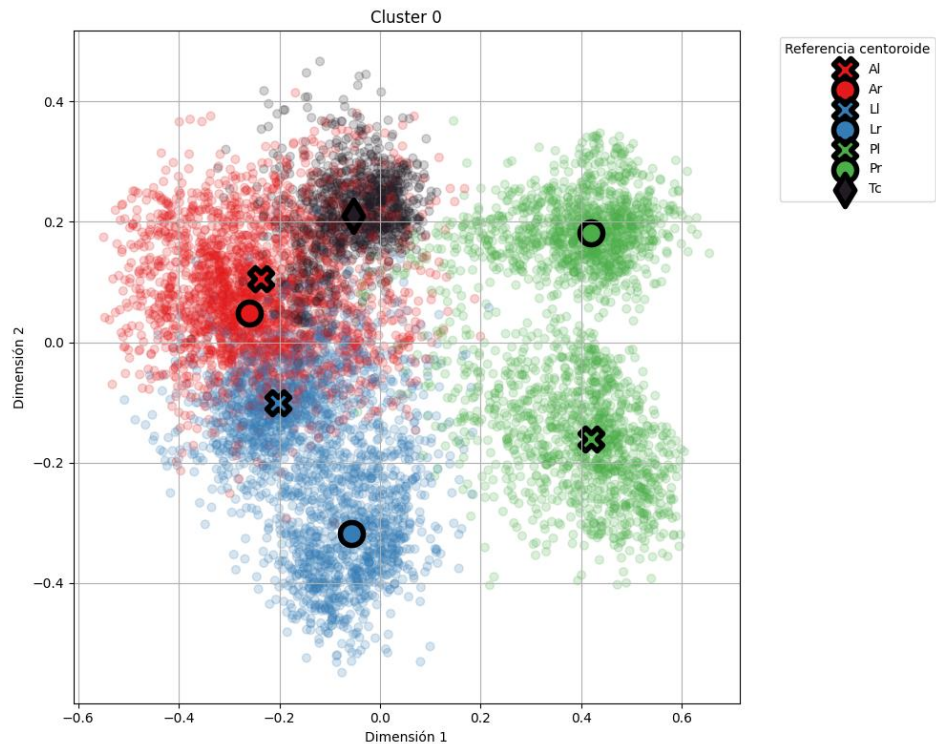


Figura 22: Muestras de clúster 0 alineadas

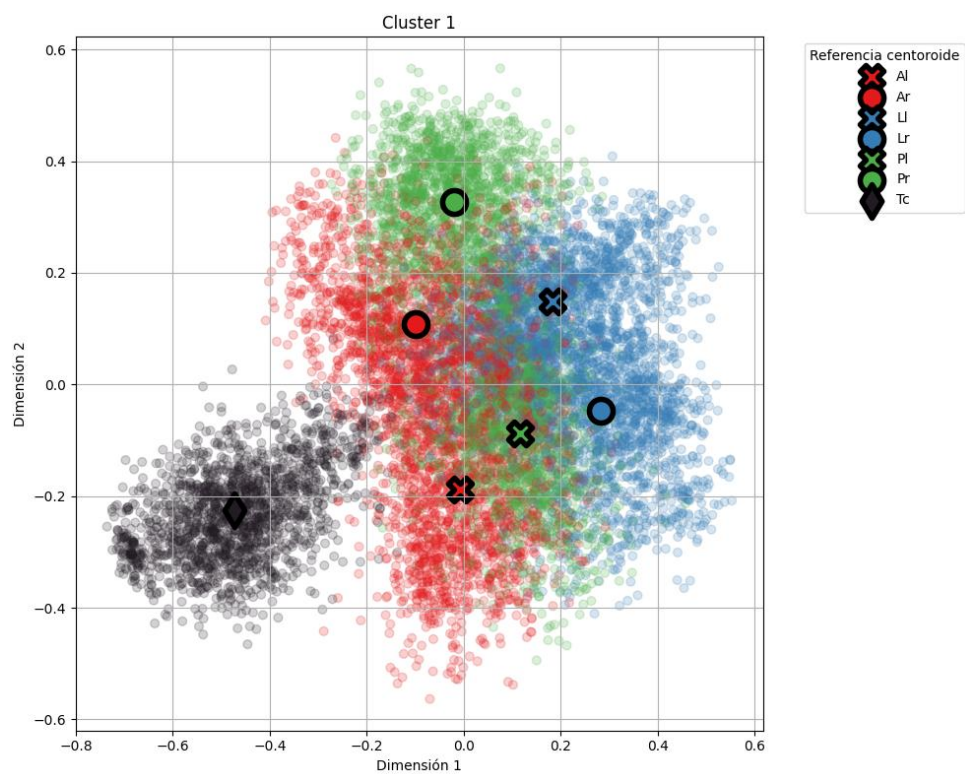


Figura 23: Muestras de clúster 1 alineadas

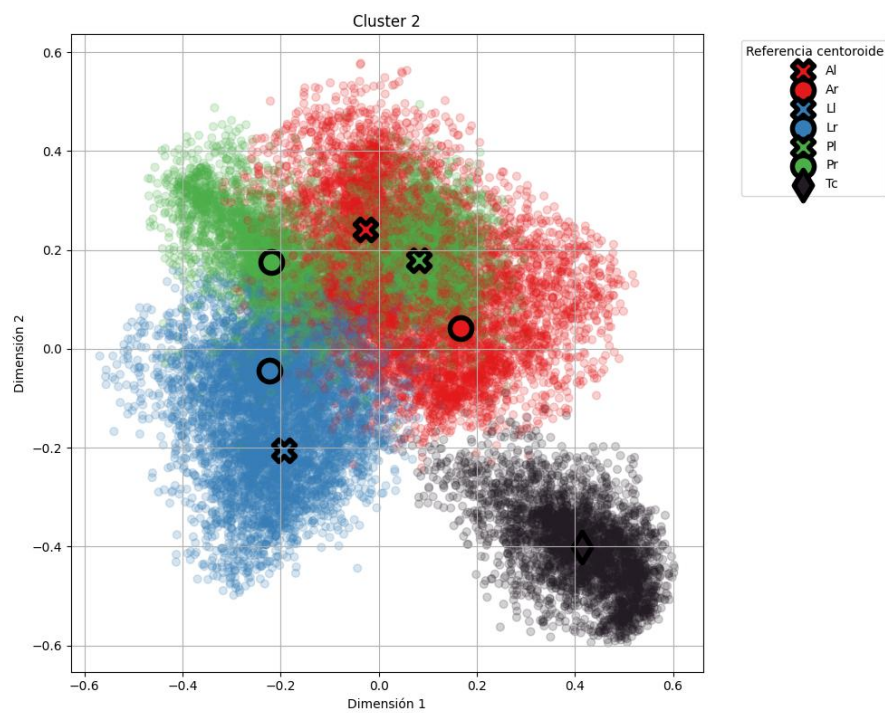


Figura 24: Muestras de clúster 2 alineadas

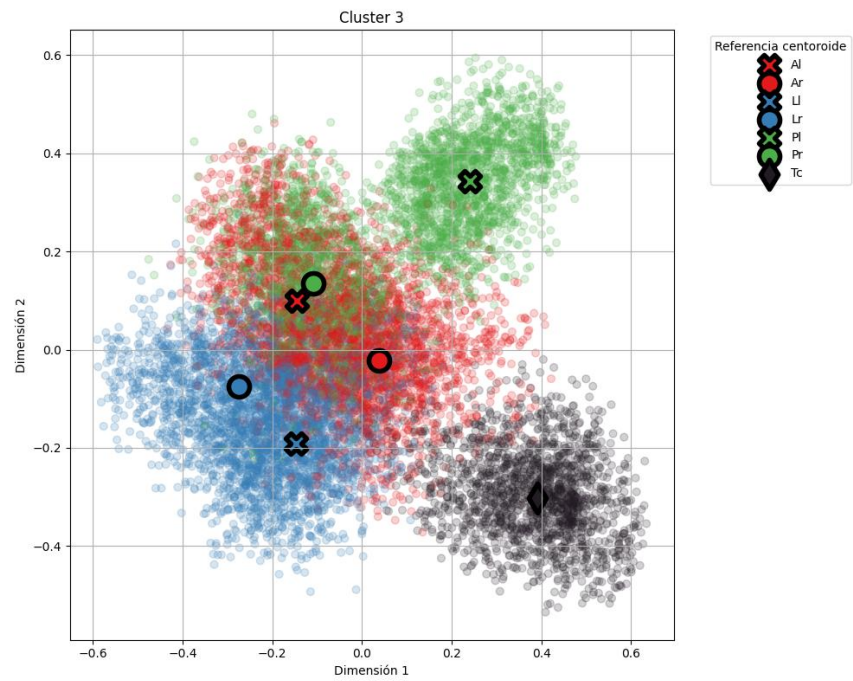


Figura 25: Muestras de clúster 3 alineadas

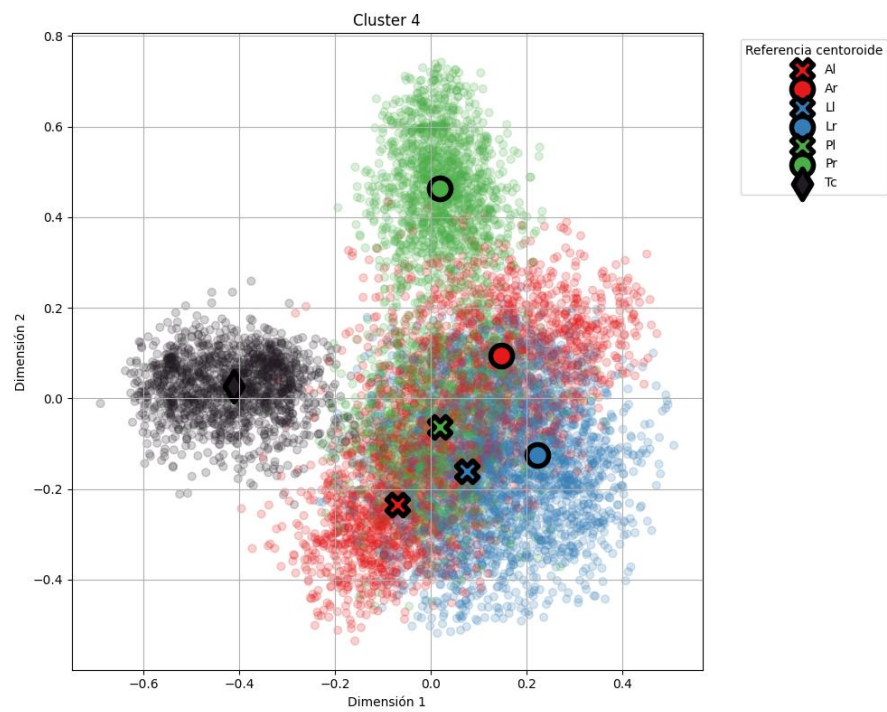


Figura 26: Muestras de clúster 4 alineadas

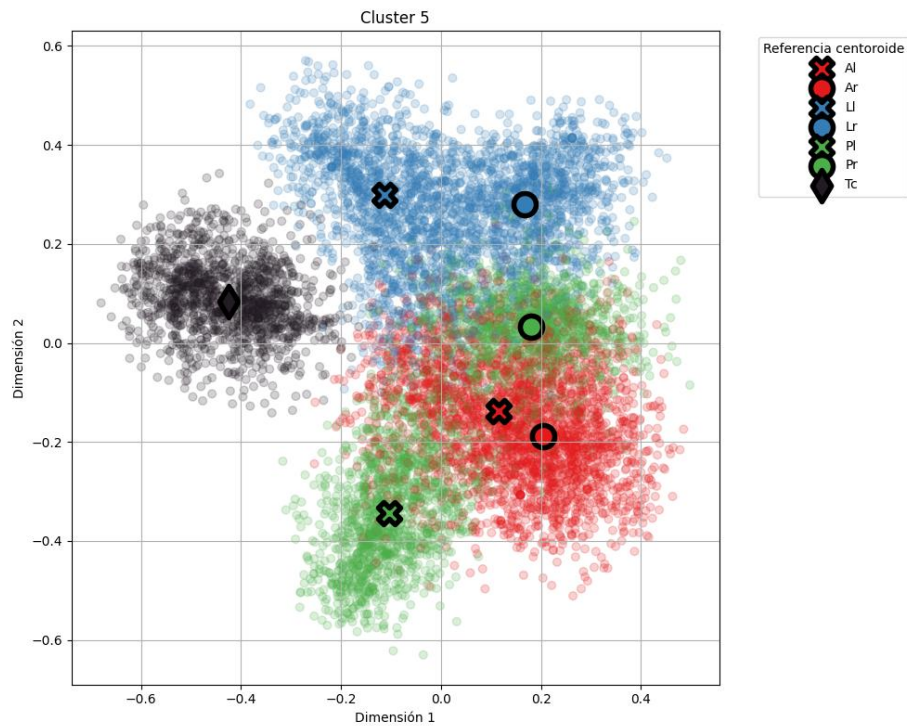


Figura 27: Muestras de clúster 5 alineadas

En la tabla 5 se puede ver el tamaño y que tan compacto es cada clúster. Para los grupos resultantes, la cantidad de segmentos asignados por clúster no es homogénea pero tampoco desbalanceada. Aquí destaca el clúster 5 llevándose más de un 30% de los segmentos. En cuanto a la compacidad de los clusteres, todos tienen una distancia promedio menor a **0.25**, por lo cual las alineaciones son estables.

Tabla 5: Tamaño y compactación de cada clúster

Clúster	Cantidad de Segmentos	Distancia Procrustes Promedio
C0	1167	0.1736
C1	1671	0.241
C2	3237	0.1927
C3	1713	0.2239
C4	1416	0.2436
C5	1420	0.203

Caracterización de eventos por clúster

En la figura 28 se muestra un promedio de número de eventos por cada 100 segmentos, esto se hace por clúster y por tipo de evento. La diferencia más notable es el número de wheezes del clúster 2, teniendo casi el doble de este tipo de eventos respecto a los otros.

Referente a la distribución de estos eventos por cada región torácica, en la figura 29 se tiene también el promedio de eventos por cada 100 segmentos, pero esta vez distribuido por las regiones torácicas. Lo primero que destaca es que, en ambos tipos de eventos, la región **LI** es la que más eventos reporta. Otro patrón destacable es el que ocurre para los wheezes en el clúster 4, la repartición es casi homogénea. Pasa algo similar para los crackles en el clúster 0.

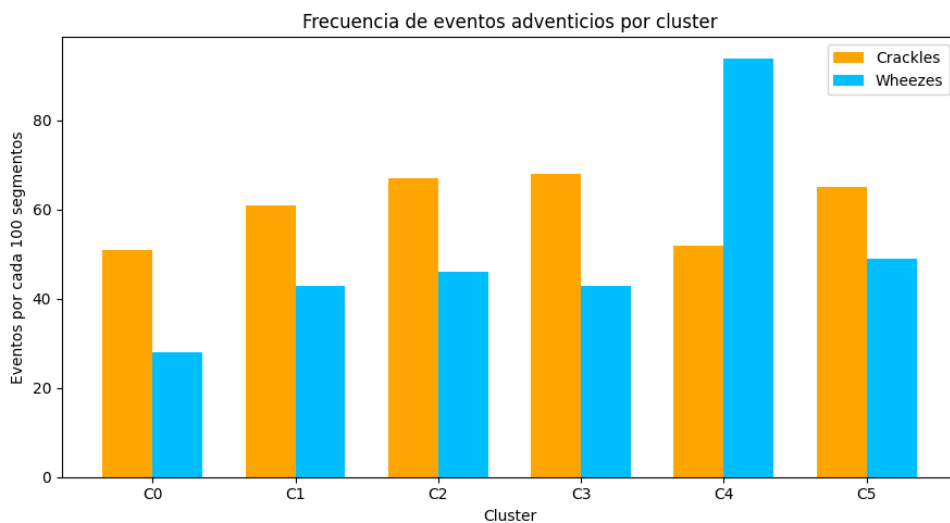


Figura 28: Distribución de eventos por clúster

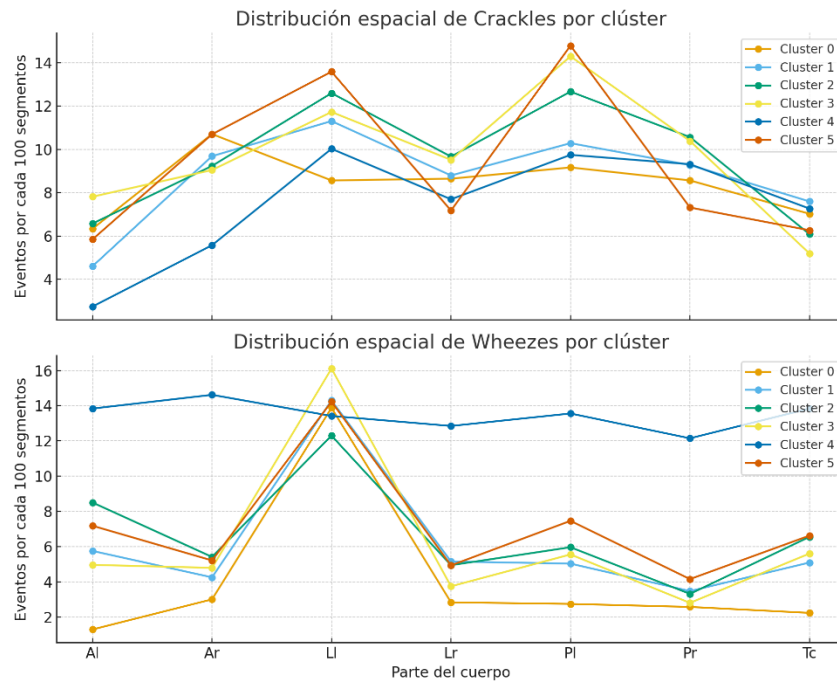


Figura 29: Distribución de eventos por clúster y región torácica

MODELADO SUPERVISADO

Construcción del conjunto de datos supervisado

Para este experimento se usan 10624 segmentos multicanal de 224 variables (7x32). Cada observación es etiquetada de forma binaria para indicar la presencia o ausencia de eventos adventicios (*wheezes* y *crackles*), resultando en la siguiente distribución de clases:

- **Crackles:** 2,342 observaciones
- **Wheezes:** 2,359 observaciones
- **Ninguno:** 5,923 observaciones

Esta distribución evidencia un desbalance, con predominio de segmentos sin eventos, lo que motiva el uso de estrategias de ponderación durante el entrenamiento.

Ajuste de hiperparámetros y entrenamiento

Se emplea un modelo de **Random Forest**, ajustando sus hiperparámetros mediante una búsqueda en rejilla con validación cruzada estratificada. El objetivo es encontrar una configuración robusta y adecuada al desbalance de clases.

Los mejores hiperparámetros obtenidos fueron:

```
{'class_weight': 'balanced', 'max_depth': 30, 'max_features': 'sqrt',  
'min_samples_leaf': 5, 'min_samples_split': 2, 'n_estimators': 300}
```

Métricas de desempeño

Tabla 6: Métricas de desempeño para el modelo global

Clase	Precision	Recall	F1-Score
Crackles	0.62	0.48	0.54
Wheezes	0.78	0.78	0.78
Ninguno	0.76	0.83	0.79
Macro promedio	0.72	0.7	0.71
Ponderado	0.74	0.74	0.74

El rendimiento es especialmente destacado en la detección de wheezes, mientras que crackles presentan mayor dificultad, en línea con su menor duración y complejidad espectral.

Importancia de características

A continuación, se evalúa la importancia acumulada y promedio de cada canal en la predicción en la tabla 7. Esto permite observar que los canales más importantes fueron Tc y Ar, esto contrasta con que estos canales son los que menos eventos tienen. Además, se analiza la importancia promedio de las características en la tabla 8 y con gran diferencia la entropía en la alta frecuencia y el primer coeficiente cepstral de frecuencia de mel, destacan como los descriptores más aportantes. El hecho de que el descriptor con mayor importancia sea en alta frecuencia viene de la mano con que en las regiones anteriores y en la tráquea, se perciben más frecuencias altas. La entropía en alta frecuencia mide que tan aleatorio es el espectro en la banda definida de 800 – 1800 Hz, mientras que el MFCC 1, es un valor relacionado con la forma general del espectro. Que estas 2 características sean las más importantes indica que los eventos estudiados tienen diferencias importantes en su perfil espectral.

Para cerrar esta parte del análisis, en la figura 30 se grafican las 20 características_regiones más importantes para el modelo, la más importante es la entropía en alta frecuencia de las señales tomadas en la parte anterior derecha del tórax, después sigue esta misma en el lateral izquierdo, zona en donde más wheezes se presentan. Además de estas 2 variables, aparecen la entropía en cada banda espectral y en diferentes regiones del tórax, lo mismo ocurre con el coeficiente cepstral 1.

Tabla 7: Importancia por cada región torácica para el modelo global

Canal	Importancia Total	Importancia Promedio
Tc	0.192	0.006
Ar	0.191	0.006
Lr	0.153	0.0048
LI	0.132	0.0041
Pr	0.117	0.0037
Al	0.112	0.0035
PI	0.103	0.0032

Tabla 8: Top 10 de importancia promedio de descriptores para el modelo global

Descriptor	Importancia Promedio
high_entropy	0.0166
mfcc_1	0.0138
mid_entropy	0.0091
low_entropy	0.0085
mid_flatness	0.0063
low_slope	0.0059
low_flux	0.0058
low_flatness	0.0056
mfcc_2	0.0054
mfcc_3	0.0049

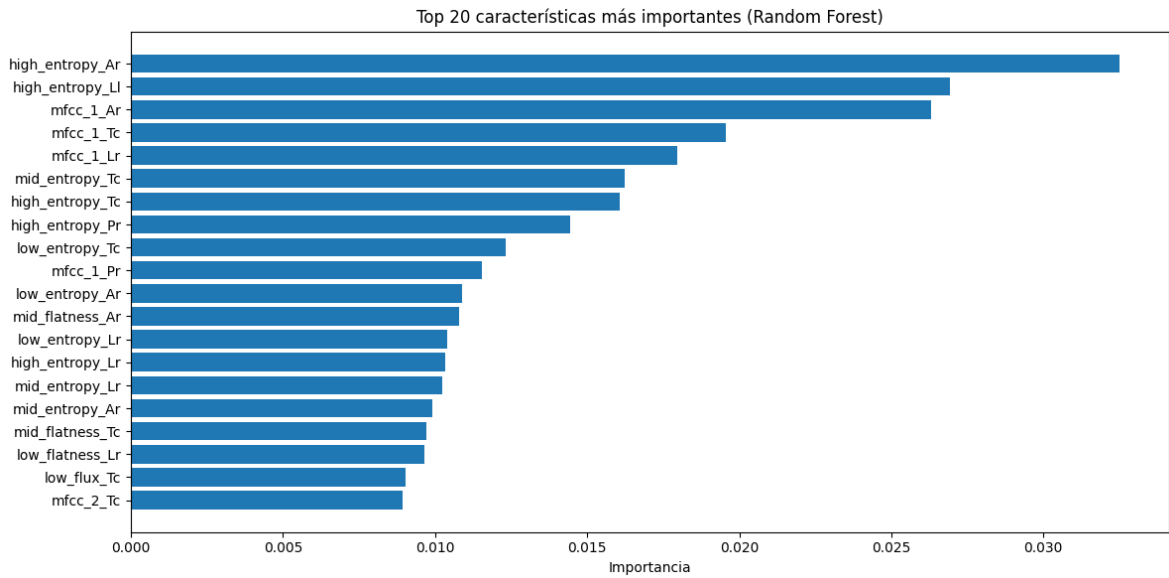


Figura 30: Importancia de características tipo descriptor_canal

Evaluación de la contribución individual de cada canal

Para evaluar la relevancia de cada canal corporal en el desempeño del modelo de clasificación, se usa una estrategia basada en la exclusión artificial de información contextual. En particular, se entrenan dos versiones del modelo para cada canal: una que incorpora todas las características disponibles y otra en la que las características del canal evaluado son anuladas (reemplazadas por ceros). Esta comparación se realiza exclusivamente sobre muestras en las que **el canal en cuestión no presenta eventos acústicos (wheezes o crackles)**, con el objetivo de observar si su presencia aporta información contextual útil para la clasificación de eventos en otras regiones.

La siguiente tabla resume la diferencia de desempeño (medida como variación absoluta en el macro F1-score) entre los modelos con y sin cada canal:

Tabla 9: Diferencias en rendimiento para pruebas de ablación

Canal	Δ Accuracy	Δ F1 Crackle	Δ F1 Wheezes
Tc	-0.01	-0.02	-0.01
Ar	0	-0.01	-0.01

LI	0	0.01	-0.01
Lr	0	-0.01	0
PI	0	0.02	0
Pr	-0.01	-0.02	0
AI	-0.01	-0.01	-0.01

En general las diferencias porcentuales son muy leves, entre 1% y 2%, pero se cómo datos relevantes, en Tc y Pr, la predicción de crackles empeora eliminando datos de estos canales, aunque no esten etiquetados con eventos. Otro dato interesante es la subida en PI para la métrica de crackles.

INTEGRACIÓN DEL ANÁLISIS SUPERVISADO Y NO SUPERVISADO

Para evaluar la utilidad práctica de la agrupación no supervisada previa, se exploran 2 estrategias, evaluar el modelo global en cada clúster y entrenar modelos supervisados específicos por clúster. Esto permite examinar si los agrupamientos reflejan divisiones reales en los datos que pueden ser aprovechadas para mejorar el rendimiento o la interpretabilidad de los modelos.

Rendimiento del modelo global por clúster

Para este experimento se seleccionan 6 conjuntos de prueba a partir de los datos del modelo global, separándolos según los clústeres definidos previamente. En la tabla 10 se presentan los resultados de estas evaluaciones. Lo más destacable es el buen desempeño en la predicción de la clase wheezes. Mientras que en la evaluación global se obtiene un F1 de 0.78, algunos clústeres logran superar esta cifra: por ejemplo, los clústeres 2 y 5 alcanzan un F1 de 0.81, y el clúster 1 logra 0.78. Esto indica que existen configuraciones espaciales donde la identificación de esta clase resulta más precisa, probablemente debido a patrones acústicos más homogéneos o representativos.

Posteriormente, se realiza un análisis SHAP para identificar las variables más influyentes en la predicción de wheezes. La tabla 11 resume el top 5 de descriptores por clúster. En esta se observa que cada grupo utiliza un conjunto diferente de características para identificar esta clase. Por ejemplo, aunque los clústeres 2 y 5 tienen el mismo valor F1, los descriptores que dominan en cada caso son distintos. El clúster 2 prioriza mfcc_1_Ar, mfcc_1_Lr y high_entropy_Tc, mientras que el clúster 5 lo hace con mfcc_1_Lr, low_slope_Lr y high_entropy_Tc. Esta variabilidad sugiere que existen múltiples rutas de predicción igualmente efectivas, adaptadas a los patrones acústicos particulares de cada grupo. Por otra parte, en el clúster 4,

destaca que las 3 variables más importante son de Pr, esto concuerda con el resultado de la figura 26, en donde esta región torácica es el punto más distante.

Tabla 10: Rendimiento del modelo global en cada clúster

Clúster	Accuracy	F1 Macro	F1 Crackles	F1 Wheezes	F1 None
0	0.74	0.68	0.53	0.72	0.81
1	0.75	0.72	0.58	0.78	0.79
2	0.75	0.70	0.49	0.81	0.80
3	0.74	0.72	0.62	0.75	0.79
4	0.74	0.68	0.45	0.76	0.81
5	0.74	0.72	0.57	0.81	0.78

Tabla 11: Características más importantes para la predicción de wheezes en el modelo global por clúster

Clúster	Top 5 Características
0	high_entropy_Tc, mid_entropy_Tc, high_entropy_Pr, mid_flatness_Tc, mfcc_1_Pr
1	mfcc_1_Lr, mfcc_1_Ar, mid_entropy_Tc, low_entropy_Ar, mid_flatness_Ar
2	mfcc_1_Ar, mfcc_1_Lr, mid_entropy_Tc, high_entropy_Tc, high_entropy_Pr
3	high_entropy_Tc, high_entropy_Ll, mfcc_1_Tc, mid_entropy_Tc, mid_flatness_Tc
4	mfcc_1_Pr, mid_entropy_Pr, high_entropy_Pr, mfcc_1_Lr, mfcc_1_Ar
5	mfcc_1_Lr, low_slope_Lr, high_entropy_Tc, low_entropy_Lr, low_flux_Lr

Entrenamiento de los modelos por clúster

En la Tabla 12, los modelos entrenados individualmente para cada clúster exhiben diferencias claras en su rendimiento. Aunque todos mantienen una precisión moderada, con accuracy entre 0.67 y 0.70, los valores de F1 macro oscilan entre 0.63 y 0.67, lo que refleja un desempeño equilibrado pero variable entre clases. El modelo global continúa superando a los modelos por clúster (F1 macro = 0.71), pero algunos submodelos, como el del clúster 5 (F1 macro = 0.67), se acercan notablemente. Esto respalda la idea de que la segmentación produce agrupaciones

con cierta coherencia interna desde el punto de vista acústico y de comportamiento del modelo. Asimismo, los valores específicos de F1 por clase indican que los modelos clústerizados tienden a balancear mejor la detección entre eventos, sin depender excesivamente de la clase dominante (*none*), aunque con un sacrificio leve en rendimiento global

Tabla 12: Rendimiento de los modelos entrenados por cada clúster

Clúster	Accuracy	F1 Macro	F1 Crackles	F1 Wheezes	F1 None
0	0.68	0.63	0.48	0.64	0.76
1	0.68	0.6	0.37	0.68	0.75
2	0.69	0.65	0.53	0.67	0.75
3	0.67	0.65	0.59	0.64	0.72
4	0.68	0.61	0.36	0.72	0.75
5	0.7	0.67	0.5	0.77	0.75
Global	0.75	0.71	0.54	0.78	0.79

El análisis de los descriptores revela una estructura altamente consistente entre clústeres. En particular, *high_entropy* y *mfcc_1* destacan como los descriptores más importantes en casi todos los modelos, en línea con los hallazgos del modelo global. Esta recurrencia sugiere que ambos capturan propiedades fundamentales de la señal acústica, probablemente relacionadas con la energía espectral y la forma general del espectro, que son claves para distinguir entre clases respiratorias.

La Tabla 13 profundiza este patrón: en todos los clústeres se observa la presencia recurrente de *high_entropy_Ar*, *mfcc_1_Ar*, *mfcc_1_Lr* o sus variantes localizadas. Además, algunos canales como *Ll* o *Tc* también aparecen repetidamente asociados a estos descriptores, lo que refuerza la idea de su relevancia funcional en múltiples configuraciones corporales.

Una observación destacada se da en el clúster 0, donde *low_slope_Lr* y *low_entropy_Lr* surgen como las dos características más relevantes, desplazando incluso a los descriptores dominantes en los otros clústeres. Esta particularidad sugiere una firma acústica específica asociada al canal derecho inferior (*Lr*) en ese grupo. El descriptor *low_slope_Lr*, en especial, no es prominente en los demás clústeres ni en el modelo global, lo que refuerza la hipótesis de que en este subconjunto de datos se manifiestan patrones únicos posiblemente relacionados con la geometría corporal revelada por el análisis de configuración. En conjunto, estos hallazgos subrayan cómo la importancia de cada descriptor no solo depende de su contenido acústico sino también del patrón espacial al que pertenece la muestra.

Tabla 13: Top 5 características para cada modelo de cada clúster

Clúster	Top 5 Características
0	low_slope_Lr, low_entropy_Lr, mfcc_1_Ar, mfcc_1_Lr, high_entropy_Pr
1	high_entropy_Ar, mfcc_1_Ar, mfcc_1_Lr, high_entropy_Pr, high_centroid_Ll
2	high_entropy_Ar, mfcc_1_Ar, high_entropy_Ll, mfcc_1_Tc, mfcc_1_Lr
3	high_entropy_Ar, mfcc_1_Ar, high_entropy_Ll, mfcc_1_Tc, high_entropy_Tc
4	mfcc_1_Ar, high_entropy_Ar, high_entropy_Tc, mfcc_1_Tc, mfcc_8_Tc
5	high_entropy_Ar, high_entropy_Ll, mfcc_1_Ar, high_entropy_Pr, mfcc_1_Lr

CONCLUSIONES

Este trabajo permitió implementar una estrategia integral de análisis estructural de sonidos respiratorios multicanal, logrando capturar no solo la información espectral local de cada posición torácica, sino también la interacción geométrica entre regiones del tórax. La construcción de la base de datos multicanal y su caracterización espectral demostró que la distribución de eventos no es homogénea: mientras L1 concentró una mayor ocurrencia de eventos adventicios, especialmente wheezes.

La aplicación de técnicas no supervisadas permitió descubrir patrones espaciales significativos en la organización de los eventos respiratorios. El uso del alineamiento geométrico y el posterior desarrollo de una clase de ProClustering basada en distancia de Procrustes representó un aporte metodológico propio, permitiendo obtener centroides espaciales interpretables y detectar configuraciones corporales recurrentes. Esto permitió pasar de un enfoque global a un análisis segmentado por patrones estructurales del tórax. Entre los resultados más notables, se identificó un clúster que mostró una distribución espacial homogénea de eventos y concentró aproximadamente el doble de wheezes respecto a los demás, sugiriendo una correlación entre configuración postural-acústica y tipo de evento.

En el análisis supervisado, el modelo Random Forest alcanzó un desempeño robusto en la detección de wheezes y patrones normales. El análisis de importancia de variables reveló sistemáticamente que Tc y Ar son regiones críticas para el proceso discriminativo, mientras que los descriptores más relevantes —como High Entropy y MFCC 1— sugieren que el contenido de alta frecuencia y la complejidad espectral son características distintivas para la clasificación. Las pruebas de ablación mostraron que eliminar un canal afecta mínimamente el desempeño global, pero sí demuestra que crackles están más distribuidos espacialmente, mientras que wheezes presentan dependencias regionales más claras.

En la etapa de integración final, se realizaron dos experimentos. En el primero, se tomó el modelo global original y se evaluó su rendimiento en subconjuntos definidos por clústeres. Este enfoque permitió aislar cómo se comporta el modelo general en configuraciones estructurales distintas. Se observó que para wheezes, el rendimiento se mantuvo o incluso superó el global (alcanzando F1 de 0.81 en clústeres 2 y 5), indicando que, en determinadas configuraciones posturales, el modelo discrimina mejor esta clase. Finalmente, un análisis SHAP permitió identificar qué características usó el modelo para predecir wheezes en cada clúster, y se evidenció que cada grupo recurrió a combinaciones diferentes de descriptor-región, por ejemplo en el clúster 4, donde en el top 3 de descriptores más importantes apareció Pr, reforzando la idea de que no solo el espectro importa, sino dónde ocurre.

En el segundo experimento, se entrenó un modelo por cada clúster, observando que en configuraciones espaciales particulares (como los clústeres 0 y 1), el rendimiento para wheezes se mantuvo relativamente alto ($F1 > 0.7$), mientras que para crackles fue significativamente menor. Además, se observan diferencias claras en la importancia de las características entre clústeres: aunque `high_entropy` y `mfcc_1` siguen siendo dominantes de forma transversal, en el clúster 0 aparece `low_slope_Lr` como descriptor principal, indicando un comportamiento acústico particular en Lr.

En conjunto, los resultados obtenidos cumplen con los objetivos planteados: se identificaron patrones espaciales en la distribución de sonidos respiratorios, se obtuvo evidencia de que los canales torácicos contienen información complementaria, y se logró integrar análisis espectral con alineamiento geométrico e interpretabilidad multicanal. Más allá del rendimiento clasificatorio, este trabajo aporta comprensión estructural sobre cómo se manifiestan los eventos acústicos en el tórax.

Como línea evolutiva natural de este estudio, será relevante explorar modelos de aprendizaje profundo estructural que no solo procesen características por canal, sino que incorporen explícitamente las relaciones espaciales entre posiciones torácicas. Se podrían desarrollar arquitecturas que operen sobre grafos (Graph Neural Networks) [39], donde los nodos representen posiciones torácicas y las aristas modelen proximidad anatómica o distancia de Procrustes. Alternativamente, podrían aplicarse modelos convolucionales sobre representaciones 2D pseudoanatómicas del tórax, o bien modelos recurrentes sobre trayectorias espacio-acústicas[38].

REFERENCIAS

- [1] X. Chen, Y. Li, and H. Liu, "Classification of respiratory sounds using optimized S-transform and deep ResNets," *Biomedical Signal Processing and Control*, vol. 72, p. 103284, 2022, doi: 10.1016/j.bspc.2022.103284.
- [2] S. Emrani, T. Gentimis, and H. Krim, "Persistent homology of delay embeddings and its application to wheeze detection," *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 459–463, 2014, doi: 10.1109/LSP.2014.2305700.
- [3] L. Fraiwan, A. Fraiwan, and H. Fardoun, "Recognition of pulmonary diseases from lung sounds using CNN and LSTM," *Biomedical Signal Processing and Control*, vol. 71, p. 103184, 2022, doi: 10.1016/j.bspc.2022.103184.
- [4] N. S. Haider and A. Behera, "Computerized lung sound-based classification of asthma and chronic obstructive pulmonary disease (COPD)," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 42–59, 2022, doi: 10.1016/j.bbe.2021.12.004.
- [5] H. Hsu et al., "Benchmarking recurrent neural networks for breath phase and adventitious sound detection," *IEEE Access*, vol. 9, pp. 60 150–60 159, 2021, doi: 10.1109/ACCESS.2021.3073409.
- [6] S. Jung et al., "Classification of lung sounds through DS-CNN models with fused STFT and MFCC features," *IEEE Transactions on Biomedical Engineering*, vol. 68, pp. 3484–3495, 2021, doi: 10.1109/TBME.2021.3067145.
- [7] P. Kapetanidis, F. Kalioras, C. Tsakonas, P. Tzamalidis, and T. G. Stavropoulos, "Respiratory diseases diagnosis using audio analysis and artificial intelligence: A systematic review," *Sensors*, vol. 24, no. 4, p. 1173, 2024, doi: 10.3390/s24041173.
- [8] Y. Kim, Y. Hyon, S. Lee et al., "The coming era of a new auscultation system for analyzing respiratory sounds," *BMC Pulmonary Medicine*, vol. 22, p. 119, 2022, doi: 10.1186/s12890-022-01896-1.

- [9] Y. Kim, Y. K. Hyon, S. S. Jung, S. Lee, G. Yoo, and C. Chung, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Scientific Reports*, vol. 11, art. no. 96724, 2021, doi: 10.1038/s41598-021-96724-7.
- [10] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.
- [11] J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964.
- [12] A. Lerch, *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Hoboken, NJ, USA: John Wiley & Sons, 2012, doi: 10.1002/9781118393550.
- [13] L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice, Eds., *Advances in Morphometrics*. New York, NY, USA: Springer, 1996.
- [14] A. Marques and C. Jácome, "Future prospects for respiratory sound research," in *Breath Sounds*. Cham, Switzerland: Springer International Publishing, 2018, pp. 291–304, doi: 10.1007/978-3-319-71824-8_16.
- [15] W. Meng et al., "Detection of respiratory sounds based on wavelet coefficients and machine learning," *Biomedical Signal Processing and Control*, vol. 50, pp. 115–123, 2019, doi: 10.1016/j.bspc.2019.101028.
- [16] A. Monaco, N. Amoroso, L. Bellantuono, E. Pantaleo, S. Tangaro, and R. Bellotti, "Multi-time-scale features for accurate respiratory sound classification," *Applied Sciences*, vol. 10, no. 23, art. no. 8606, 2020, doi: 10.3390/app10238606.
- [17] J. A. Muñoz Barraza, C. N. Saavedra Benardis, R. S. Adasme Jeria, D. H. Arellano Sepúlveda, I. Sánchez Díaz, and P. J. Bertrand Navarrete, "Concordance in discriminating recordings of different lung sounds between physiotherapists," *Respiratory Care*, vol. 65, no. 2, pp. 177–182, 2020, doi: 10.4187/respcare.06624.

- [18] F. G. Nabi, K. Sundaraj, C. K. Lam, and R. Palaniappan, "Characterization and classification of asthmatic wheeze sounds according to severity level using spectral integrated features," *Computerized Biology and Medicine*, vol. 104, pp. 52–61, 2019, doi: 10.1016/j.compbimed.2018.10.035.
- [19] H. Pasterkamp, P. L. Brand, M. Everard, L. Garcia-Marcos, H. Melbye, and K. N. Priftis, "Towards the standardisation of lung sound nomenclature," *European Respiratory Journal*, vol. 47, no. 3, pp. 724–732, 2016, doi: 10.1183/13993003.01132-2015.
- [20] F. J. Rohlf and D. Slice, "Extensions of the Procrustes method for the optimal superimposition of landmarks," *Systematic Zoology*, vol. 39, no. 1, pp. 40–59, 1990.
- [21] M. Sarkar, I. Madabhavi, N. Niranjana, and M. Dogra, "Auscultation of the respiratory system," *Annals of Thoracic Medicine*, vol. 10, no. 3, pp. 158–168, 2015, doi: 10.4103/1817-1737.160831.
- [22] Sociedad Chilena de Pediatría, "Recomendaciones para el uso del estetoscopio en pediatría," *Revista Chilena de Pediatría*, vol. 91, no. 4, pp. 500–505, 2020, doi: 10.4067/S0370-41062020000400500.
- [23] R. N. Shepard, "The analysis of proximities: Multidimensional scaling with an unknown distance function," *Psychometrika*, vol. 27, no. 2, pp. 125–140, 1962.
- [24] S. Singh, "Respiratory symptoms and signs," *Medicine*, 2020, doi: 10.1016/j.mpmed.2020.01.001. (nota: completa volumen/número si los usas en el texto)
- [25] A. Sovijärvi, F. Dalmaso, J. Vanderschoot, L. Malmberg, G. Righini, and S. Stoneman, "Definition of terms for applications of respiratory sounds," *European Respiratory Review*, vol. 10, no. 77, pp. 597–610, 2000.
- [26] W. S. Torgerson, *Theory and Methods of Scaling*. New York, NY, USA: Wiley, 1958.

- [27] B. Yesilkaya, M. Perc, and Y. Isler, "Manifold learning methods for the diagnosis of ovarian cancer," *Journal of Computational Science*, vol. 63, art. no. 101775, 2022, doi: 10.1016/j.jocs.2022.101775.
- [28] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
- [31] C. Molnar, *Interpretable Machine Learning*. Independently published, 2022.
- [32] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [33] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
- [34] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinformatics*, vol. 8, p. 25, 2007.
- [35] A. Jones, R. D. Jones, K. Kwong, and Y. Burns, "Effect of positioning on recorded lung sound intensities in subjects without pulmonary dysfunction," *Phys. Ther.*, vol. 79, no. 7, pp. 682–690, 1999, doi: 10.1093/ptj/79.7.682.
- [36] L. Rodríguez-Giraldo and J. Ugarte, "Correlating EEG Signals and Electrode Locations by Means of Multidimensional Scaling," in *Proc. 16th Int. Joint Conf. on Biomedical Engineering Systems and Technologies (BIOSTEC 2023)*, vol. 4: BIOSIGNALS, 2023, pp. 309–316, doi: 10.5220/0011750200003414.

- [37] S. Escobar and J. Ugarte, "Auscultation chest locations are correlated with short-term spectral features of pulmonary sounds," in Proc. Int. Conf. on Mathematical Analysis and Applications in Science and Engineering – Book of Extended Abstracts, ISEP / P.PORTO, Jun. 2022.
- [38] J. Yu, C. Zhang, and S. Wang, "Multichannel one-dimensional convolutional neural network-based feature learning for fault diagnosis of industrial processes," *Neural Comput. Appl.*, vol. 33, pp. 3085–3104, 2021, doi: 10.1007/s00521-020-05171-4.
- [39] J. Zhou et al., "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2021, doi: 10.1016/j.aiopen.2021.01.001.
- [40] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability, vol. 1, 1967, pp. 281–297.
- [41] J. C. Gower, "Generalized Procrustes Analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975, doi: 10.1007/BF02291478.
- [42] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. Chichester, U.K.: John Wiley & Sons, 1998.
- [43] A. Andreella and L. Finos, "Procrustes Analysis for High-Dimensional Data," *Psychometrika*, vol. 87, pp. 1422–1438, 2022, doi: 10.1007/s11336-022-09859-5.

ANEXOS

A. ORIGEN DE LOS DATOS

Los audios, archivos de eventos y archivos de metadatos utilizados en este proyecto están disponibles en:

https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge

B. SCRIPTS UTILIZADOS

Los scripts, funciones y clases utilizadas en este proyecto se encuentran disponibles en el repositorio de github:

<https://github.com/SesbastianEP29/An-lisis-Respiratorio-con-Procrustes>

Aquí están las funciones para la lectura y creación de los conjuntos de datos:

- Assign_events_to_windows
- Extract_metadata
- process_audio
- spectral_descriptors
- extract_features_from_segments

La función del algoritmo de referencia centroide y la clase proclustering:

- procrustes_align_to_centroid
- proclusters

Y modelos mejores modelos de RF resultantes:

- Modelo_global.pkl
- Modelos de las pruebas de ablación: Modelo_Tc.pkl, Modelo_Ar.pkl...etc.
- Modelos por clúster: Modelo_c0.pkl, Modelo_c1.pkl... etc