

Universidad de La Habana  
Facultad de Matemática y Computación



# **Análisis Computacional de Coros de *Eleutherodactylus eileenae*: Inferencia de Redes desde Señales Bioacústicas**

**Autor: Daniel Machado Pérez**

**Tutores:**

**Dr. Roberto Mulet Genicio**

**Dr. Milton García Borroto**

Trabajo de Diploma  
presentado en opción al título de  
Licenciado en Ciencia de la Computación



La Habana,  
mayo 2025

[github.com/DanielMPMatCom/Thesis.git](https://github.com/DanielMPMatCom/Thesis.git)

*"La utopía está en el horizonte  
Camino dos pasos y se aleja dos pasos.  
Camino diez y el horizonte se corre diez pasos más allá.  
Entonces, ¿para qué sirve la utopía?  
Para eso, sirve para caminar."*

**Fernando Birri**

# Agradecimientos

El autor agradece al Dr. Roberto Alonso Bosch y la M.Sc. Mariam Curbelo Cruz por facilitar los datos de campo, compartir su amplia experiencia sobre el comportamiento de *Eleutherodactylus eileenae*, orientar la revisión de la literatura especializada y la revisión del manuscrito.

# Resumen

El análisis computacional de coros animales representa un desafío en ecoacústica, particularmente en ambientes con superposición de señales y ruido. Esta investigación desarrolló un flujo automatizado para procesar grabaciones de *Eleutherodactylus eileenae*, una rana endémica de Cuba cuyos machos forman coros nocturnos para atraer a las hembras. Se aplicó un filtrado espectral basado en percentiles para atenuar interferencias de banda ancha y puntuales. A continuación, nueve pistas de audio se sincronizaron mediante correlación cruzada. Se implementaron dos algoritmos heurísticos para detectar y asignar cantos a micrófonos: uno basado en hipótesis de energías relativas a los audios y otro mediante *clustering* espectro-temporal, este último mostrando mayor precisión en condiciones de baja actividad. Las series temporales resultantes se modelaron como una red de espines usando el modelo de Ising, cuyos parámetros  $J_{ij}$  se inferieron por máxima verosimilitud. Aunque se identificaron acoplamientos significativos, el modelo independiente superó al de Ising en la predicción de patrones, sugiriendo interacciones débiles o limitaciones en la formulación estándar del modelo. El trabajo demuestra que enfoques heurísticos reproducibles permiten reconstruir dinámicas colectivas en coros y propone mejoras metodológicas para futuros estudios.

# Abstract

The computational analysis of animal choruses poses a significant challenge in ecoacoustics, particularly in environments with signal overlap and noise. This study developed an automated pipeline to process recordings of *Eleutherodactylus eileenae*, a frog endemic to Cuba whose males form nocturnal choruses to attract females. First, spectral filtering based on amplitude percentiles was applied to attenuate both broadband and transient interference. Then, nine audio tracks were synchronized via cross-correlation. Two heuristic algorithms were implemented to detect and assign calls to microphones: one based on hypotheses of relative energy in the recordings and another using spectro-temporal clustering, the latter demonstrating higher accuracy under low activity conditions. The resulting time series were modeled as a spin network using the Ising model, with interaction parameters  $J_{ij}$  inferred via maximum likelihood. Although significant couplings were identified, an independent model outperformed the Ising model in predicting call patterns, suggesting either weak interactions or limitations in the standard model formulation. This work demonstrates that reproducible heuristic approaches can reconstruct collective dynamics in animal choruses and proposes methodological improvements for future studies.

# Índice general

<b>Introducción</b>	<b>10</b>
0.1. Motivación y justificación . . . . .	11
0.2. Formulación del problema . . . . .	12
0.3. Hipótesis . . . . .	14
0.4. Objetivos . . . . .	14
0.5. Propuesta de solución . . . . .	15
0.6. Estructura del Documento . . . . .	15
<b>1. Antecedentes</b>	<b>17</b>
1.1. Detección de individuos en grabaciones . . . . .	18
1.1.1. Métodos clásicos de clasificación de señales acústicas	18
1.1.2. Localización y asignación de fuentes . . . . .	19
1.2. Análisis de coros y comportamiento grupal . . . . .	19
1.2.1. Patrones de formación de coros . . . . .	20
1.2.2. Ecoacústica de coros . . . . .	20
1.3. Modelos de interacción y causalidad . . . . .	21
1.3.1. Modelos estadísticos de redes . . . . .	21
1.3.2. Técnicas de inferencia de causalidad . . . . .	22
1.4. Antecedentes específicos en <i>E. eileenae</i> . . . . .	22
<b>2. Métodos</b>	<b>24</b>
2.1. Representación de audios con Mel-espectrogramas . . . . .	26
2.2. Filtrado de ruido de fondo . . . . .	27
2.3. Sincronización de los audios . . . . .	28
2.4. Detección y asignación heurística de cantos de Colines a Mi- crófonos . . . . .	30
2.4.1. Algoritmo basado en hipótesis de energías relativas .	30
2.4.2. Algoritmo basado en agrupamiento espectro-temporal	32
2.5. Modelado de interacciones con el modelo de Ising . . . . .	33

2.5.1. Fundamentos del modelo de Ising . . . . .	34
2.5.2. Formulación del sistema acústico como red de espines	34
2.5.3. Inferencia por máxima verosimilitud y descenso por gradiente . . . . .	35
<b>3. Resultados e Interpretación</b>	<b>38</b>
3.1. Resultados y evaluación de los algoritmos propuestos . . . .	38
3.1.1. Consistencia . . . . .	39
3.1.2. Comparación con detecciones manuales . . . . .	40
3.2. Comparación entre algoritmos . . . . .	43
3.3. Análisis de las interacciones inferidas $\{J_{ij}\}$ . . . . .	46
3.4. Análisis de idoneidad del modelo de Ising . . . . .	49
<b>Conclusiones</b>	<b>51</b>
<b>Recomendaciones</b>	<b>53</b>
<b>Referencias Bibliográficas</b>	<b>54</b>

# Índice de figuras

1.	Ejemplar de <i>Eleutherodactylus eileenae</i> . . . . .	11
2.	Distribución Geográfica de los Micrófonos. . . . .	13
2.1.	Mel-espectrograma de un fragmento del audio 20231021_190000h, donde se observan las notas “CO” (baja frecuencia) y “LIN” (alta frecuencia). . . . .	27
2.2.	Mel-espectrograma del fragmento del audio 20231021_190000b, en el que aparece el ruido de un motor de automóvil. Antes del filtrado. . . . .	28
2.3.	Mel-espectrograma del fragmento del audio 20231021_190000b, en el que se registró el ruido de un motor de automóvil. Después del filtrado. . . . .	29
3.1.	Fragmento de Secuencia de Cantos de los Colines en los archivos de 20231021_190000. Arriba la señales “LIN”, debajo los “CO”. Cada color representa una rana o micrófono diferente. . . . .	39
3.2.	Matriz de correlación entre ejecuciones independientes del algoritmo sobre los datos de 20231021_190000. Los bloques diagonales de alta correlación (marcados en rojo) muestran la consistencia de las detecciones para cada micrófono. . . .	41
3.3.	Matriz de confusión del algoritmo de energías relativas vs. detecciones manuales (izq.). Matriz de confusión del algoritmo de <i>clustering</i> espectro-temporal vs. detecciones manuales (der.). Solo se consideraron los cantos detectados en las grabaciones con mayor actividad. . . . .	42
3.4.	<i>Boxplot</i> y <i>violinplot</i> de $ \Delta t $ entre “CO” y “LIN” para detecciones manuales, algoritmo de energías relativas y <i>clustering</i> . .	42



3.5. Histogramas de $\Delta t$ : (izq.) manual vs. energías relativas; (der.) manual vs. <i>clustering</i> . Los colores representan el origen de los datos (etiquetado manual, algoritmo de energías relativas o algoritmo de <i>clustering</i> ). . . . .	43
3.6. Comparación entre los tiempos de detección de cantos por los algoritmos A (energías relativas) y B (comportamiento de la especie) sobre los datos de 20231021_190000. . . . .	44
3.7. Histogramas comparativos de $\Delta t$ entre "CO" y "LIN" para el Algoritmo A (energías temporales) y el Algoritmo B ( <i>clustering</i> de espectro-tiempo). . . . .	45
3.8. Matriz de interacciones $J_{ij}$ inferida para el 20231021_190000. . . . .	47
3.9. Grafo de interacciones correspondiente al 20231021_190000. . . . .	48
3.10. Evolución de la matriz $J_{ij}$ durante tres horas consecutivas (20231021_180000, 20231021_190000, 20231021_200000). . . . .	48
3.11. Evolución temporal de los grafos de interacciones inferidas para tres horas consecutivas (20231021_180000, 20231021_190000, 20231021_200000). . . . .	49
3.12. Tasa de aparición de patrones de canto observada versus predicha por el modelo independiente (azul) y el modelo de Ising (naranja) para las grabaciones de 20231021_190000 (izq.) y 20231021_040000 (der.). . . . .	50

# Introducción

Las técnicas de monitorización acústica automatizada se han convertido en una herramienta esencial para el estudio de comunidades de animales en su medio natural. En particular, el análisis del canto de los anuros ofrece información valiosa sobre sus estrategias de apareamiento, la dinámica de sus poblaciones y la salud de los ecosistemas en los que habitan [5]. La rana cubana *Eleutherodactylus eileenae* (Dunn, 1929), endémica de Cuba, constituye un modelo idóneo para investigar estos procesos, pues sus machos forman coros nocturnos cuyos patrones de interacción pueden reflejar aspectos tanto ecológicos como de comportamiento social [2].

*E. eileenae*, conocido comúnmente como “Colín”, es una especie perteneciente a la familia *Eleutherodactylidae*. Su distribución abarca desde la península de Guanahacabibes, en Pinar del Río, hasta la Sierra de Najasa, en Camagüey, ocupando diversos hábitats como bosques semideciduos, pinares y cafetales, hasta altitudes de aproximadamente 900 metros sobre el nivel del mar [2, 37, 17]. Durante el día, estos anuros se refugian en oquedades de rocas, troncos, hojarasca y bromelias; al anochecer, los machos ascienden a ramas y hojas, hasta alturas de 3 metros, para emitir sus vocalizaciones.

El canto característico de *E. eileenae* consta de dos notas diferenciadas: una primera “Co”, breve y de frecuencia menor, seguida de una segunda “Lin”, más prolongada y de frecuencia ligeramente superior. Esta estructura ha motivado el uso del término “Colín” para referirse tanto a cada canto individual como, coloquialmente, a la especie misma. Estas llamadas desempeñan un papel crucial en la comunicación intraespecífica, mediando aspectos relacionados con el estado fisiológico y la competencia territorial [2]. Además, se ha observado que los machos forman agrupaciones o “coros”, sincronizando sus cantos para aumentar la eficacia en la atracción de hembras y reducir el riesgo de depredación.

La presente investigación desarrolla un flujo computacional automatizado para procesar grabaciones de los cantos de los Colines y, a partir

de él, comenzar a analizar las interacciones acústicas asociadas a estos coros. El objetivo es comprender las estrategias de comunicación que emplea *E. eileenae* en su entorno natural. En la Figura 1 se muestra un ejemplar de esta especie, y en la Guía sonora de los anfibios de Cuba <sup>1</sup> se puede encontrar una breve grabación de su canto y otros datos relevantes.



Figura 1: Ejemplar de *Eleutherodactylus eileenae*.

## 0.1. Motivación y justificación

La ecoacústica ha evolucionado desde la grabación pasiva hasta flujos computacionales que automatizan la detección de llamadas en ambientes ruidosos [1, 5]. Sin embargo, como resultado de la revisión bibliográfica realizada, no se conocen estudios en Cuba que hayan integrado estas técnicas con modelos de interacción y causalidad.

---

<sup>1</sup>Grabación disponible en: [https://www.fonozoo.com/fnz\\_detalle\\_registro\\_amphibia.php?id=97953&tipo\\_registro=1](https://www.fonozoo.com/fnz_detalle_registro_amphibia.php?id=97953&tipo_registro=1)

El análisis manual de grabaciones de campo resulta lento, costoso y propenso a errores humanos, lo que limita la escala y reproducibilidad de los estudios bioacústicos. En trabajos previos sobre *E. eileenae*, los investigadores describieron manualmente las etapas de vocalización y realizaron muestreos puntuales de la actividad acústica y trófica de los machos en la Sierra del Rosario [2]. Sin embargo, no se abordó la sincronización automática de múltiples micrófonos ni la clasificación sistemática de cada llamado, pasos indispensables para escalar el análisis a largas series temporales y diferentes localidades.

Además, los coros de machos implican interacciones acústicas cuya dinámica no se comprende completamente: ¿cómo influyen la densidad de individuos, la configuración espacial de los micrófonos y el ruido ambiental en la estructura del coro? ¿Qué reglas simples de interacción subyacen en la sincronización de los “Colines”? En este contexto, un enfoque computacional reproducible y basado en algoritmos heurísticos permitiría procesar de manera eficiente decenas de horas de grabaciones, discriminando eventos de interés (cantos de Colines) de otros sonidos (otra fauna, viento, tráfico) y garantizando resultados comparables entre campañas de muestreo.

Los resultados de la presente tesis constituyen un primer paso hacia el desarrollo de un flujo completo que abarque, desde la adquisición y sincronización de señales, hasta la clasificación automática de cantos y el modelado de interacciones acústicas.

## 0.2. Formulación del problema

El presente trabajo se ocupa del procesamiento automático de un conjunto de grabaciones registradas en la Reserva de la Biosfera “Sierra del Rosario” con el objetivo de caracterizar los cantos de machos de *Eleutherodactylus eileenae* y reconstruir la dinámica de sus coros. Estas grabaciones, realizadas con nueve micrófonos omnidireccionales dispuestos en un área de aproximadamente 20 m de radio, abarcan ciclos de 58 minutos de registro seguidos de 2 minutos de descarga, durante tres noches consecutivas entre las 18:00 y las 06:00 horas. El entorno de captura se ve afectado por ruido de lluvia, viento, tráfico y la superposición de cantos de varios individuos, lo que plantea retos importantes tanto en la sincronización de señales como en la discriminación de los eventos acústicos relevantes frente a emisiones no deseadas. En la Figura 2 se muestra la distribución geográfica de los micrófonos en el área de estudio.

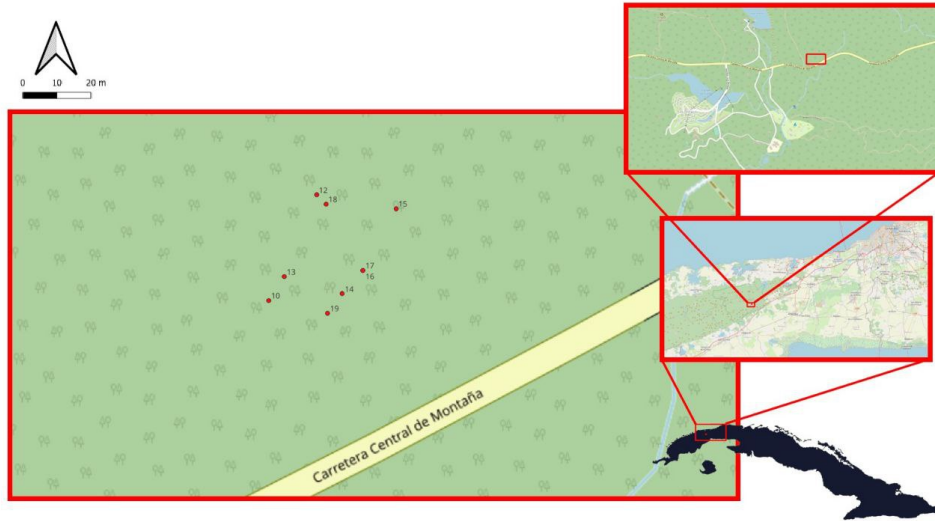


Figura 2: Distribución Geográfica de los Micrófonos.

En este ámbito, el problema central consiste en el diseño de un flujo heurístico capaz de sincronizar con precisión las nueve pistas de audio, detectar de manera fiable los instantes de emisión de cada canto (“Colín”) y asignar cada evento al micrófono (y por tanto al individuo) correspondiente; a partir de ello, se extraen parámetros temporales y espectrales de cada llamado para generar un registro estructurado que sirve de base a la inferencia de una red de interacción inspirada en el modelo de Ising [23]. Se espera que la sincronización automática alcance una resolución temporal tal que permita distinguir llamadas superpuestas de machos vecinos, y que la combinación de criterios de proximidad y similitud espectral garantice una alta exactitud en la etiqueta de cada canto, validada mediante comparaciones de rendimientos y muestreo manual. Finalmente, las secuencias temporales resultantes se emplean para estimar los parámetros de interacción  $I_{ij}$ , de modo que el modelo reproduzca con fidelidad las correlaciones de co-emisión observadas en el coro.

La complejidad del problema reside en la superposición de señales y en la variabilidad del nivel de ruido ambiental, lo cual exige emplear técnicas robustas de filtrado. Este trabajo opta por heurísticas de intensidad relativa y correlación espectral para asignar cada “Colín” al micrófono y al individuo correspondiente, garantizando la reproducibilidad de los resultados

y la comparabilidad con futuros estudios bioacústicos. Este estudio es el primero en Cuba en aplicar el modelo de Ising para describir interacciones acústicas en coros de anuros, siguiendo aproximaciones similares en redes neuronales y colonias de insectos [35, 27, 4, 33, 23].

### 0.3. Hipótesis

Se plantea como hipótesis principal que un flujo computacional basado en algoritmos heurísticos para la sincronización, detección y clasificación de cantos de Colines podrá reproducir las secuencias de emisión de cada individuo y, a partir de ellas, inferir las interacciones acústicas subyacentes a la formación de coros.

Esto fundamenta la viabilidad de un enfoque computacional replicable y escalable que, aplicado a los cantos de Colines, aporte nuevas perspectivas cuantitativas sobre la dinámica de los coros y sirva como base para futuros estudios comparativos en ecoacústica.

### 0.4. Objetivos

El objetivo general de esta tesis es desarrollar y validar un flujo heurístico que, mediante la detección y clasificación de los cantos de *E. eileenae*, permita reconstruir la dinámica de sus coros y cuantificar su sincronización y causalidad. Para ello se plantean los siguientes objetivos específicos:

#### 1. Construir un *dataset* limpio y sincronizado:

- Sincronizar las nueve pistas de audio utilizando referencias comunes (picos de energía) y correlación cruzada. [14]
- Eliminar ruidos de fondo (lluvia, viento, tráfico) e interferencias no deseadas.

#### 2. Detectar y asignar cada canto a su emisor:

- Desarrollar heurísticas de proximidad basadas en la intensidad relativa en cada micrófono.
- Validar la asignación mediante muestreo manual.
- Evaluar la consistencia de los algoritmos de detección y asignación.
- Comparar el rendimiento de los algoritmos diseñados.

### 3. Inferir un modelo de interacciones acústicas:

- Formular la red de interacciones como un modelo de Ising con parámetros  $J_{ij}$  que cuantifiquen la co-emisión.
- Estimar los  $J_{ij}$  mediante el Principio de Máxima Verosimilitud [19] y algoritmos de descenso por gradiente.
- Evaluar la idoneidad del modelo de Ising mediante una comparación con el modelo Independiente en cuanto a su capacidad para predecir los patrones de cantos.

## 0.5. Propuesta de solución

Para dar respuesta al problema planteado, se propone un flujo de trabajo modular que integre cuatro etapas: sincronización, filtrado, asignación de emisores y modelado de interacciones.

La sincronización de las nueve pistas de audio se realiza mediante la aplicación de correlación cruzada sobre los histogramas de distribución de distancias temporales entre los picos de energía de los audios y un punto de referencia. A continuación, se aplica un filtro que consiste en una poda espectral basada en el percentil 99.9 de la amplitud por frecuencia. Con ello se busca atenuar ruido de baja (viento, tráfico) y alta frecuencia (insectos).

En la tercera fase se asigna cada llamado al micrófono más adecuado mediante heurísticas de intensidad relativa. Finalmente, la información temporal y espectral extraída de cada evento sirve para construir un grafo de interacciones, modelado con un enfoque de máxima verosimilitud en un sistema análogo al modelo de Ising. El ajuste de los parámetros  $J_{ij}$  se lleva a cabo mediante un algoritmo de descenso por gradiente.

Este flujo, además de automatizar completamente la extracción de datos, permite validar la hipótesis mediante métricas objetivas de precisión y reproducibilidad, y ofrece una herramienta extensible a otros sistemas de ecoacústica con configuraciones similares.

## 0.6. Estructura del Documento

Para explicar la solución propuesta y el flujo de trabajo, se diseñó un documento con la siguiente estructura. En el Capítulo 1 se revisan los antecedentes y actualidad de la detección y clasificación de llamadas bioacústicas, con especial énfasis en métodos de sincronización y filtrado de señales.

El Capítulo 2 detalla la metodología empleada: adquisición y preprocesamiento de las grabaciones, tanto el diseño de algoritmos heurísticos para la detección de los “Colines” como la asignación de cada canto a su emisor, y la formulación del modelo de interacciones basado en Ising con inferencia por máxima verosimilitud y descenso por gradiente. En el Capítulo 3 se presentan los resultados: evaluación de la precisión de sincronización, rendimiento de los detectores de canto, exactitud de la asignación micrófono-Colín y análisis de las interacciones inferidas  $\{J_{ij}\}$ . Además en las Conclusiones se discuten los aportes principales, las implicaciones ecológicas y computacionales y las limitaciones del estudio.

Finalmente, en las Recomendaciones se presentarán sugerencias generales basadas en los resultados, destinadas a orientar mejoras metodológicas, proponer nuevas líneas de investigación y facilitar la aplicación práctica del enfoque desarrollado en estudios futuros.



# Capítulo 1

## Antecedentes

Las primeras grabaciones sistemáticas de sonidos de vida silvestre se realizaron en mayo de 1929, cuando Arthur A. Allen y Peter P. Kellogg registraron por primera vez el canto de especies neárticas en un parque de Ithaca, Nueva York [13]. Este hito marcó el nacimiento de la colección que hoy conocemos como Macaulay Library, el mayor archivo multimedia de fauna del mundo, con más de 2.6 millones de grabaciones de audio y cobertura de casi todas las especies de aves, así como registros de insectos, peces, anfibios y mamíferos [12].

Durante décadas, la bioacústica de campo dependió de hardware analógico y procesamiento manual de cintas, lo cual limitaba el volumen de datos y la reproducibilidad de los estudios [5]. Con la transición a sistemas digitales a principios de los años 2000, la monitorización acústica pasiva (PAM) emergió como herramienta clave en ecología y conservación, permitiendo grabaciones continuas y remotas que facilitan el muestreo a gran escala sin perturbar el comportamiento natural de los organismos [1].

Las organizaciones de conservación, entre ellas WWF, impulsaron la adopción de PAM en áreas protegidas, destacando la necesidad de protocolos robustos de preprocesamiento para atenuar ruido ambiental (como lluvia, viento y tráfico) y mejorar la relación señal-ruido [5]. Paralelamente, el desarrollo de índices acústicos automatizados, introducidos por Towsey et al. (2014), permitió estimar la riqueza de especies en paisajes sonoros complejos mediante métricas computables directamente de las grabaciones [44].

La creación de grandes conjuntos de datos de referencia, como AnuraSet para llamadas de anuros neotropicales, estandarizó la evaluación de algoritmos de identificación en entornos ruidosos, facilitando la compara-

ción de metodologías y la reproducibilidad de los resultados [9]. Revisiones como la de Stowell (2022) han descrito las oportunidades y desafíos de incorporar redes neuronales profundas en bioacústica, estableciendo una hoja de ruta para futuros avances computacionales [40].

En el ámbito del modelado de interacciones, Ōta et al. (2020) cuantificaron los mecanismos de sincronización en coros de ranas a partir de características dinámicas de las llamadas, demostrando la aplicabilidad de enfoques de red para estudiar la co-emisión y la causalidad acústica [29]. Asimismo, Kalra (2024) examinó la percepción de señales en ambientes complejos, proporcionando métodos de localización pasiva que mejoran la atribución de fuentes en PAM [25].

Estos avances tecnológicos y metodológicos forman el fundamento de la presente investigación, que adapta flujos computacionales y modelos estadísticos de interacción al estudio de los coros de *Eleutherodactylus eileenae* en su entorno natural cubano.

## 1.1. Detección de individuos en grabaciones

La detección precisa de emisores individuales en grabaciones de campo constituye un paso esencial para cualquier estudio bioacústico que aspire a caracterizar patrones de comportamiento a nivel de individuos. Este proceso implica la identificación de eventos acústicos relevantes en señales, la discriminación de llamadas frente a ruido ambiental y, cuando es posible, la atribución de cada emisión a su emisor de origen. A continuación, se revisan primero los métodos clásicos de clasificación de señales basados en análisis espectral y extracción de características, y luego se exponen enfoques de localización pasiva y heurísticas de asignación que permiten estimar la procedencia de las llamadas.

### 1.1.1. Métodos clásicos de clasificación de señales acústicas

La clasificación de vocalizaciones animales en grabaciones se fundamenta en la extracción de características espectrales a partir de la Transformada Rápida de Fourier (FFT) [11]. Sobre esta base, la conversión a coeficientes cepstrales en frecuencia mel (MFCC) ha demostrado ser especialmente eficaz para representar texturas acústicas complejas, tal como lo estableció Mermelstein (1976) en el ámbito de voz humana y luego adaptado a bioacústica [26]. Zhang et al. (2021) emplearon la descomposición de espectrogramas mel y la fusión de modelos para clasificar escenas acústi-

cas, alcanzando mejoras significativas en la precisión de clasificación [50]. Xie et al. (2016) extendieron este enfoque integrando detección de eventos acústicos con aprendizaje multi-etiqueta sobre espectrogramas, lo que permitió discriminar actividades de canto de ruido de lluvia con resultados prometedores [47]. Estos métodos clásicos forman la columna vertebral de los flujos heurísticos de detección, siendo especialmente adecuados cuando solo se dispone de una única pista de audio.

### 1.1.2. Localización y asignación de fuentes

Aunque la verdadera localización tridimensional de emisores requiere múltiples canales sincronizados, diversos autores han explorado técnicas de triangulación y diferencias de tiempo de llegada (TDOA) para estimar direcciones de origen. Spiesberger y Fristrup (1990) detallaron un método de TDOA que calcula desplazamientos de fase entre pares de micrófonos para inferir la dirección de la fuente sonora [38]. Cuando solo se cuenta con grabaciones monofónicas, se han diseñado heurísticas basadas en la intensidad relativa de la señal y la similitud espectral. Kalra (2024) demostró que, al comparar la energía de una llamada en momentos consecutivos y su correlación espectral con plantillas predefinidas, es posible asignar eventos a individuos cercanos con una precisión alta [25]. Estas heurísticas aprovechan la atenuación del sonido en el espacio y la coherencia espectral para discriminar entre vocalizaciones de distintos machos dentro de una misma grabación, constituyendo una alternativa pragmática cuando no se dispone de redes de micrófonos múltiples.

## 1.2. Análisis de coros y comportamiento grupal

El estudio de los coros de anuros combina observaciones de campo con análisis cuantitativos para desentrañar cómo las interacciones individuales dan lugar a patrones colectivos de canto. Este enfoque implica medir la distribución espacial y temporal de las vocalizaciones, identificar picos de actividad sincronizada y describir las reglas de proximidad y fase que los machos siguen al unirse al coro. A continuación se revisan los principales hallazgos sobre la formación de coros y las herramientas ecoacústicas desarrolladas para caracterizar su estructura y dinámica.

### 1.2.1. Patrones de formación de coros

En *Eleutherodactylus coqui*, Woolbright (1985) registró que los individuos ascienden al anochecer a perchas de la vegetación, donde emiten series regulares de “co-qui” alternadas con períodos de silencio, y observó que la densidad de cantores influye en la cercanía entre perchas, reduciéndose la distancia media en noches de alta actividad [46]. Stewart y Pough (1983) encontraron en la misma especie que la localización de los machos en el espacio de canto está relacionada con la disponibilidad de refugios y la competencia territorial, ajustando su posición en respuesta a la presencia de vecinos [39]. Drewry y Rand (1983) describieron patrones similares en *E. coqui* en Puerto Rico, donde la sincronización de picos vocales ocurre en respuesta a variaciones de luz y temperatura, con crestas al anochecer y al amanecer [15].

En el caso de *Eleutherodactylus eileenae* en Cuba, Alonso et al. (2001) analizaron la actividad acústica de machos cantores y revelaron dos picos de emisión al anochecer y al amanecer, así como desplazamientos verticales en la vegetación que enfatizan la dimensión espacial de la formación de coros [2]. Estos hallazgos concuerdan con los patrones descritos en otras especies de *Eleutherodactylus*, donde la competencia acústica y las condiciones microambientales determinan la estructura del coro [43]. Más recientemente, Òta et al. (2020) cuantificaron dinámicamente las interacciones entre pares de individuos mediante análisis de series temporales de llamadas, demostrando que algunos machos actúan como “iniciadores” de pulsos colectivos, un comportamiento que puede describirse con modelos de red de interacción [29].

### 1.2.2. Ecoacústica de coros

La ecoacústica de coros se apoya en el uso de índices acústicos y herramientas computacionales para detectar y caracterizar la actividad colectiva sin necesidad de seguimiento visual. Boelman et al. (2007) introdujeron el Bioacoustic Index (BI), calculado como la energía acumulada en la banda de frecuencias de interés [6], mientras que Pieretti et al. (2011) complementaron este enfoque con el Acoustic Complexity Index (ACI), que mide variaciones espectrales dentro de ventanas temporales para estimar la riqueza y la intensidad coral, mostrando alta correlación con el número de machos activos en grabaciones de bosques tropicales, y demostraron que ambos índices permiten resaltar los períodos de mayor sincronización en coros de anfibios [31]. Además, Farina (2018) estableció un marco cuantitativo para

la ecoacústica que enfatiza la integración de índices acústicos en estudios de dinámica comunitaria, subrayando su utilidad para evaluar patrones de sincronización en coros de anuros y otros sistemas biológicos [18].

En 2021, Gan et al. propusieron un método de reconocimiento de coros basado en espectrogramas falsos en color y aprendizaje automático, logrando identificar patrones de emisión colectiva con una precisión del 77% en grabaciones de campo de especies tropicales [20]. Brodie et al. (2020) evaluaron la eficacia de múltiples índices acústicos y confirmaron que la combinación de ACI, BI y Net Spectrum ofrece una representación robusta de la estructura coral a lo largo de ciclos nocturnos [7].

### 1.3. Modelos de interacción y causalidad

En ciencias biológicas y ecológicas, comprender cómo las interacciones locales entre individuos dan lugar a fenómenos colectivos constituye un desafío central. Los modelos estadísticos de redes, en particular las formulaciones de máxima entropía equivalentes al modelo de Ising, ofrecen un marco potente para describir la sincronización y la causalidad en sistemas de canto animal. En esta sección se revisan primero las bases y aplicaciones de los modelos de Ising en redes biológicas, y luego se presentan las principales técnicas de inferencia de causalidad adaptadas a series temporales acústicas.

#### 1.3.1. Modelos estadísticos de redes

El modelo de Ising, originado en física estadística para describir ferromagnetismo, se reformuló como un problema de máxima entropía para ajustar distribuciones observadas de estados binarios con correlaciones dadas. Schneidman et al. (2006) demostraron que un modelo de Ising con interacciones par a par reproduce con alta fidelidad las correlaciones de actividad en poblaciones de neuronas retinianas, sin necesidad de términos de orden superior [35]. Mora y Bialek (2011) ampliaron esta perspectiva, argumentando que sistemas biológicos (desde redes neuronales hasta colonias de insectos) operan cerca de puntos críticos de fase, un rasgo que maximiza la capacidad de información y la sensibilidad a perturbaciones externas [27].

Para inferir los parámetros de interacción  $J_{ij}$  en sistemas fuera de equilibrio, Zeng et al. (2013) desarrollaron un método de máxima verosimilitud para modelos de Ising asíncronos, derivando reglas de aprendizaje que

dependen únicamente de las correlaciones temporales de los datos [49]. Chau Nguyen et al. (2017) revisaron un amplio espectro de métodos de inferencia inversa (incluyendo pseudoverosimilitud y aproximaciones de mensaje) que escalan eficientemente a grandes redes biológicas, destacando aplicaciones en neurociencia y ecología [10]. Estos enfoques han permitido reconstruir estructuras de interacción tanto en poblaciones de neuronas (Tkačik et al., 2006) como en sistemas de canto colectivo, donde las “conexiones” cuantifican la probabilidad de co-emisión entre pares de individuos [42].

### 1.3.2. Técnicas de inferencia de causalidad

Más allá de la correlación, establecer relaciones causales en series temporales acústicas requiere herramientas específicas. El enfoque clásico de Granger (1969) define la causalidad en términos predictivos: una serie  $X$  *cause*<sup>a</sup> otra  $Y$  si el pasado de  $X$  mejora la predicción de  $Y$  más allá de lo que hace el propio pasado de  $Y$  [21]. Porta y Faes (2015) ampliaron este marco en el contexto de sistemas fisiológicos y redes neuronales, describiendo cómo la causalidad de Wiener-Granger puede aplicarse a señales biológicas con metodologías que consideran tanto ruido como dinámicas no lineales [32].

Para capturar relaciones más complejas, se han empleado medidas de información direccional como la entropía de transferencia (Schreiber, 2000), que cuantifica flujos de información sin asumir linealidad [36]. Más recientemente, el método de Convergent Cross Mapping (Sugihara et al., 2012) ha mostrado robustez al inferir causalidad en sistemas dinámicos de alta dimensionalidad, al reconstruir espacios de fase conjuntos y medir la habilidad de uno de los vectores de recuperación del otro [41]. Estas técnicas, que serán utilizadas en trabajos futuros, proporcionan un conjunto de herramientas complementarias para analizar la causalidad acústica en coros de *E. eileenae*, permitiendo distinguir interacciones directas de sincronizaciones inducidas por factores ambientales comunes.

## 1.4. Antecedentes específicos en *E. eileenae*

La investigación acerca de la biología y ecología de *Eleutherodactylus eileenae* arrancó fundamentalmente con estudios taxonómicos y de distribución. Schwartz (1958) describió por primera vez la especie en la región occidental de Cuba, señalando su afinidad con el subgénero *Eleutherodacty-*

*lus auriculatus* y detallando caracteres morfológicos diagnósticos [37]. Posteriormente, Estrada (1984, 1994) amplió el conocimiento de su rango geográfico, documentando nuevas localidades en Sancti Spíritus y otros municipios centrales, así como preferencias de hábitat que incluyen bosques semideciduos, pinares y cafetales hasta los 900 m s. n. m. [16, 17]. Estas obras sentaron las bases para comprender la distribución espacial de *E. eileenae* y su adaptación a distintos sustratos, como bromelias y hojarasca.

En el plano de la bioacústica y la ecología trófica, Alonso et al. (2001) llevaron a cabo muestreos en la Reserva de la Biosfera “Sierra del Rosario”, donde cuantificaron patrones estacionales de vocalización de los machos y relacionaron la actividad acústica con la alimentación mediante análisis de contenido estomacal [2]. Encontraron dos picos de canto, al anochecer y al amanecer, coincidentes con desplazamientos verticales en la vegetación, y demostraron que durante los períodos de depresión vocal los machos se alimentan, consumiendo principalmente himenópteros y arácnidos. Este estudio fue el primero en integrar datos acústicos y tróficos de *E. eileenae*, estableciendo el marco experimental y metodológico que la presente tesis extiende mediante flujos computacionales automatizados y modelado de interacciones acústicas.

## Capítulo 2

# Métodos

El canto de *E. eileenae* ocurre durante la noche, cuando los machos se agrupan para vocalizar, con picos de actividad en los meses cálidos y lluviosos del año, y en las horas de la tarde-noche y la madrugada. Los datos empleados en esta investigación fueron obtenidos por un grupo de investigación liderado por el Dr. Roberto Alonso, de la Facultad de Biología de la Universidad de La Habana, mediante un conjunto de grabaciones de campo realizadas en la Reserva de la Biosfera “Sierra del Rosario”. Luego de identificar ciertos individuos de Colín y los sitios en los que habitualmente se ubicaban para emitir sus cantos, se colocaron nueve micrófonos (cada uno a aproximadamente un metro de su individuo más cercano) cuya activación remota permitió registrar la información acústica del coro. Todos los dispositivos, del mismo modelo y omnidireccionales, fueron activados simultáneamente para grabar durante 58 minutos consecutivos, seguidos por un intervalo de 2 minutos destinado a la descarga y almacenamiento de los datos adquiridos. Este ciclo se repitió cada hora, desde las 18:00 hasta las 06:00 del día siguiente, durante tres noches consecutivas, entre el 20 y el 23 de octubre de 2023. Como resultado, se conformó un *dataset* estructurado en el que, por cada hora de grabación, se dispone de los nueve archivos correspondientes a los nueve micrófonos utilizados. Tal como se indicó previamente, los dispositivos fueron ubicados dentro de un área de aproximadamente 20 metros de radio.

Debido a que los micrófonos eran omnidireccionales y estaban situados relativamente cercanos entre sí, en cada uno se registraron no solamente los cantos del individuo de Colín más próximo, sino también cantos de especímenes cercanos, sonidos del ambiente natural de la zona, e incluso ruidos de agentes artificiales como autos pasando por una carretera aledaña.



Esto planteó una serie de dificultades para el análisis de los datos, como la de eliminar la información ruidosa, y la de distinguir y extraer en cada audio los cantos del Colín que corresponden al micrófono en cuestión.

De esta forma, se impuso emplear un método automático para “limpiar” los audios. Además, para poder obtener las secuencias de cantos y estudiar los coros, surgió la necesidad de diseñar procesos para discriminar correctamente estos datos, y obtener una asignación Micrófono-Colín. También se debía verificar la correcta sincronización de las grabaciones, pues a pesar de la activación remota y simultánea, posibles errores de medición podrían comprometer la precisión de los resultados. En este capítulo se describen los métodos utilizados para resolver tanto estos problemas como el de modelar las interacciones del sistema.

Con el objetivo de simplificar las explicaciones, se decidió asignar una letra entre la “a” y la “i” a cada uno de los micrófonos, de modo que la correspondencia entre la notación original y la nueva queda de la siguiente forma:

- Micrófono 10: a
- Micrófono 12: b<sup>1</sup>
- Micrófono 13: c
- Micrófono 14: d
- Micrófono 15: e
- Micrófono 16: f
- Micrófono 17: g
- Micrófono 18: h
- Micrófono 19: i

Por lo tanto, una grabación en específico, por ejemplo la del 21 de octubre de 2023 a las 19:00 horas, registrada por el micrófono “h”, será referida como “20231021\_190000h”.

Para la grabación se emplearon dispositivos AudioMoth, desarrollados como placas de circuito impreso de 58x48x18mm (80g con baterías), basados en un microcontrolador ARM Cortex-M4F con unidad de punto flotante y un micrófono MEMS bottom-ported. Cada unidad muestreó audio

---

<sup>1</sup>En los datos originales no existe micrófono 11.

sin compresión WAV a velocidades de hasta 384kHz, almacenando los datos en tarjetas microSD de hasta 32GB. Gracias a su excepcional eficiencia energética (con consumos de 10-25mW durante la adquisición y 80μW en reposo), resultaron idóneos para despliegues prolongados alimentados por tres pilas AA, manteniendo la sincronización y calidad requeridas para el estudio de los coros de *E.eileenae* [22].

## 2.1. Representación de audios con Mel-espectrogramas

La escala de mel es una escala perceptual de tono que aproxima la respuesta no lineal del oído humano a diferentes frecuencias. A diferencia de la escala lineal en Hertz (Hz), enfatiza la discriminación de frecuencias bajas (e.g., 100-1000 Hz) y comprime las altas (e.g., >4000 Hz), reflejando cómo los humanos perciben diferencias de altura tonal. Ampliamente utilizada en procesamiento de audio, permite representar señales acústicas en formatos espectralmente significativos llamados Mel-espectrogramas, facilitando la detección de patrones relevantes. [50] Un Mel-espectrograma se obtiene aplicando una ventana deslizante al audio, calculando la FFT en cada segmento, mapeando las bandas de frecuencia a la escala mel y, finalmente, tomando el logaritmo de la energía resultante [50].

Para el análisis de las grabaciones se transformó cada archivo de audio en su correspondiente Mel-espectrograma, buscando enfatizar las variaciones espectrales relevantes para el canto de *E. eileenae* y facilitar la detección de los pulsos “CO” y “LIN” en los audios.

Para la generación de los Mel-espectrogramas se emplearon los siguientes parámetros, optimizados para cubrir el rango de frecuencias de los cantos de Colín (aprox. 1.8-3.1 kHz) [3] y garantizar resolución temporal suficiente para distinguir emisiones consecutivas:

- `hop_length` = 512 (desplazamiento de ventana)
- `n_fft` = 6096 (número de puntos para calcular la FFT)
- `n_mels` = 128 (número de bandas mel a utilizar)
- `f_min` = 1600 Hz (frecuencia mínima mapeada)
- `f_max` = 4096 Hz (frecuencia máxima mapeada)
- `sample_rate` = 96000 Hz (tasa de muestreo de las grabaciones)

En la Figura 2.1 se muestra un Mel-espectrograma de un fragmento del audio 20231021\_190000h. En él pueden identificarse claramente los instantes de emisión de la nota baja “CO” (fragmento en torno a la 20 banda de mel) y de la nota alta “LIN” (bandas intensas por encima de la 60).

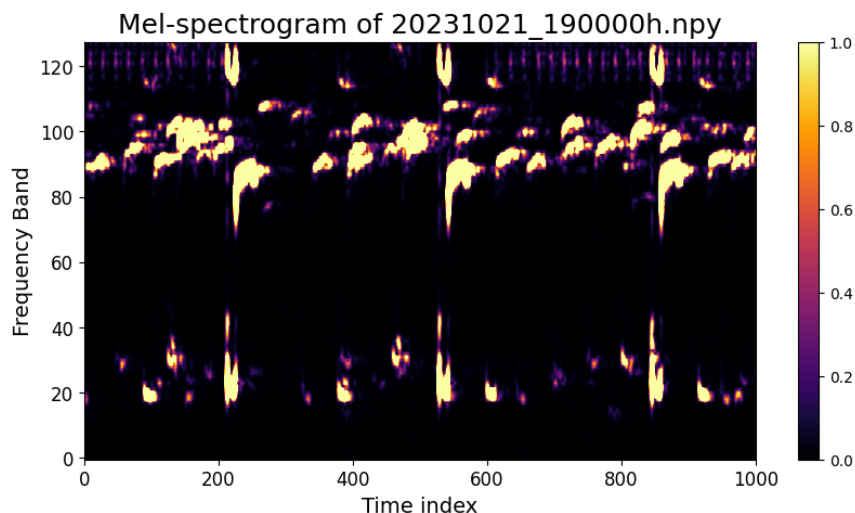


Figura 2.1: Mel-espectrograma de un fragmento del audio 20231021\_190000h, donde se observan las notas “CO” (baja frecuencia) y “LIN” (alta frecuencia).

## 2.2. Filtrado de ruido de fondo

Para garantizar que las características acústicas extraídas representaran de manera fidedigna la emisión vocal de los Colines, se implementó una etapa de eliminación de ruido dirigida a atenuar tanto interferencias puntuales (por ejemplo, vocalizaciones de otras especies o clics mecánicos) como ruidos persistentes de banda ancha (motores de vehículos, canto de grillos, viento). En la Figura 2.2 se muestra el Mel-espectrograma de un fragmento del audio 20231021\_190000b, en el que se registró el ruido generado por el paso de un automóvil cercano.

La técnica empleada consistió en una poda espectral (o *spectral pruning*) basada en el percentil 99.9 de la amplitud por frecuencia. Para cada grabación se calculó el espectrograma de potencia y, en cada banda de frecuencia, se determinó el valor correspondiente al percentil 99.9 de la amplitud a lo

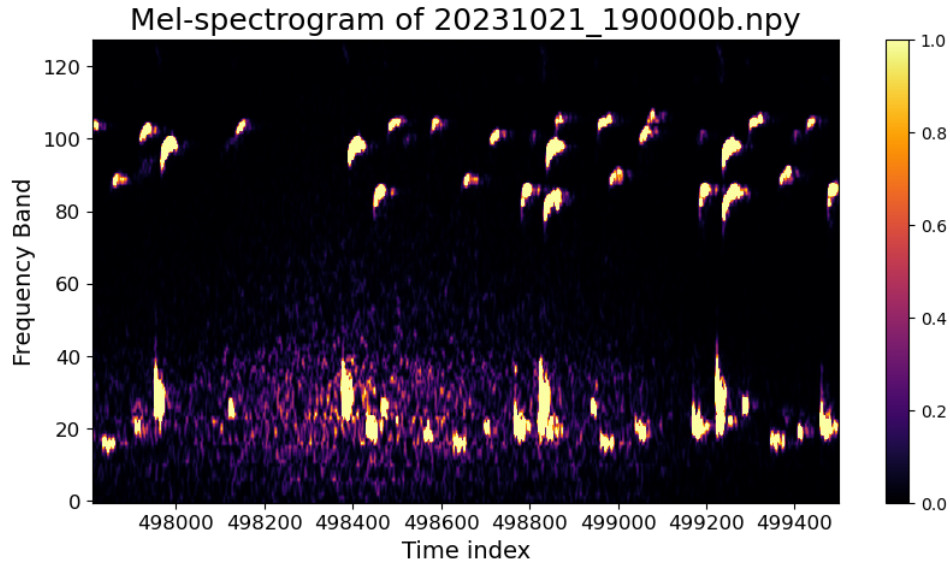


Figura 2.2: Mel-espectrograma del fragmento del audio 20231021\_190000b, en el que aparece el ruido de un motor de automóvil. Antes del filtrado.

largo de todo el período analizado. A continuación, se descartaron todas las componentes espectrales por debajo de ese umbral, reteniendo solo los eventos acústicos cuyas intensidades superaron el fondo de ruido común. En la Figura 2.3 se presenta el Mel-espectrograma del mismo fragmento de audio que se muestra en la Figura 2.2, pero luego de aplicar el filtrado.

Este procedimiento permitió suprimir eficazmente las componentes de baja energía asociadas al ruido ambiental no informativo y conservar las estructuras espectrales dominantes de los cantos de *E. eileenae*.

### 2.3. Sincronización de los audios

La sincronización de las grabaciones de audio constituyó una etapa crítica para asegurar la coherencia temporal necesaria en el análisis posterior de los cantos de *E. eileenae*. Aunque los nueve dispositivos de grabación fueron activados simultáneamente, cada uno presentaba desfases mínimos, posiblemente debido a pequeñas variaciones en el hardware y al retardo de la activación remota, los cuales resultaban significativos para el procesamiento subsecuente. Para corregir estos desfases se implementó un pro-

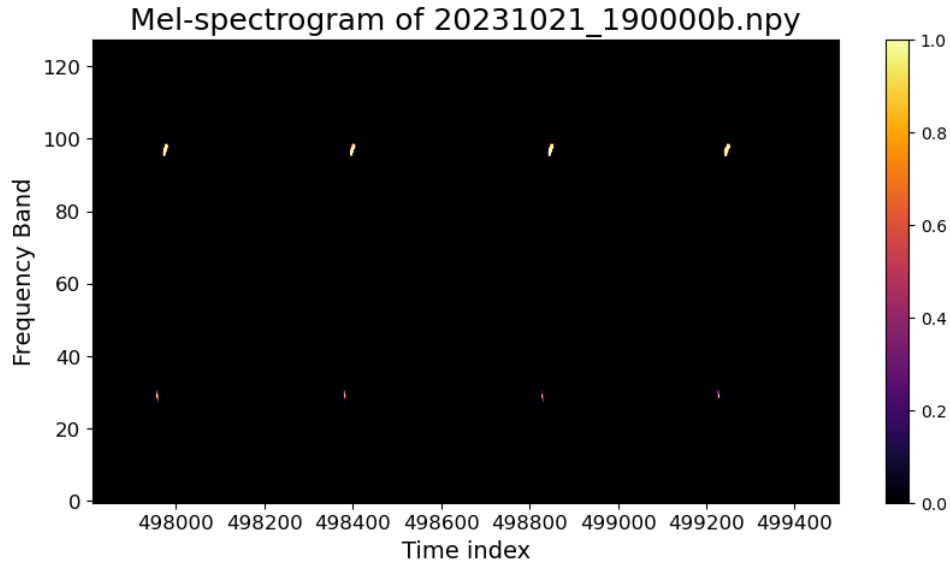


Figura 2.3: Mel-espectrograma del fragmento del audio 20231021\_190000b, en el que se registró el ruido de un motor de automóvil. Después del filtrado.

cedimiento automático basado en los puntos de máximos locales extraídos de los Mel-espectrogramas de cada canal.

En primer lugar, tras eliminar el ruido de fondo, se localizaron los puntos de máximo local en cada espectrograma. Dado que un Mel-espectrograma es una matriz de dimensiones (bandas de frecuencias x tiempo), cada punto de máximo se identifica por un par  $(t, f)$ , donde  $t$  es el índice temporal y  $f$  la banda de frecuencia asociada.

A continuación, para cada par de archivos se tomaron sus respectivas listas de puntos de máximo local y se alinearon en tiempo buscando el desplazamiento óptimo. Con tal fin, se construyeron histogramas de los índices temporales de los máximos (ignorando la componente de frecuencia) y se aplicó correlación cruzada entre ambos histogramas. La correlación cruzada es una medida estadística que cuantifica la similitud entre dos señales al compararlas en diferentes desplazamientos temporales, utilizada comúnmente para sincronizar grabaciones o identificar patrones coincidentes.

en análisis de series temporales. Se calcula con la siguiente fórmula:

$$CC = \sum_{i=0}^n a_i b_i, [14]$$

donde  $a_i$  y  $b_i$  son los valores en el índice  $i$  de los *arrays* sobre los que se aplica el método. En este caso el desplazamiento (*lag*) que maximiza la correlación indica la mejor coincidencia temporal entre los picos de energía de los dos canales.

Una vez obtenidos los *lags* óptimos para cada par de archivos, se empleó un algoritmo *greedy* para propagar estas alineaciones y establecer una sincronización global entre los nueve canales. Finalmente, los desplazamientos resultantes se aplicaron a los datos originales, garantizando que todos los audios compartieran una referencia temporal común con un error mínimo.

## 2.4. Detección y asignación heurística de cantos de Colines a Micrófonos

Para obtener un *dataset* con el que fuera posible llevar a cabo los estudios de comportamiento de los coros, se hizo imprescindible diseñar procesos automáticos capaces de extraer, a partir de las grabaciones sincronizadas y filtradas, las series temporales de emisión de cada Colín, asignando correctamente cada canto a su respectivo micrófono. Solo disponiendo de estas secuencias individuales se puede avanzar hacia un análisis riguroso de las interacciones acústicas y las dinámicas de grupo, ya que cualquier estudio de causalidad o modelado de red requiere un sistema de eventos bien definido y libre de ambigüedades en la identificación de los emisores. En esta sección se presentan dos algoritmos heurísticos que, partiendo de Mel-espectrogramas, automatizaron la detección de instantes de canto y su asignación espacial, constituyendo la base metodológica para el posterior estudio de las redes de interacción en el coro de *E. eileenae*.

### 2.4.1. Algoritmo basado en hipótesis de energías relativas

El método propuesto se sustenta en la suposición de que para un micrófono, los cantos del individuo más cercano deben ser registrados con las mayores energías relativas a dicho micrófono. Por lo tanto, si en cada uno se identifican los momentos de mayor energía y se elimina la información

de esos cantos de las demás grabaciones, se obtendrán por separado los datos que se buscan. Para explotar esta idea se dividió cada grabación en dos bandas de interés (baja frecuencia (“CO”) y alta frecuencia (“LIN”)) y se procesaron en paralelo. En cada banda se computaron las energías temporales, obteniendo en cada instante  $j$ , la suma de los cuadrados de las intensidades en las  $n$  bandas de frecuencia del Mel-espectrograma, es decir:

$$E_j = \sum_{i=0}^n a_{ij}^2 \quad \forall j, \quad (2.1)$$

donde  $a_{ij}$  es la intensidad en la banda  $i$  y el tiempo  $j$ , y  $E_j$  representa la energía total en dicho instante [24].

A partir de estos vectores de energía, el algoritmo aplicó la siguiente estrategia:

1. Se creó un *array* de ceros para cada micrófono, que almacenaría la energía del canto asignado en cada instante.
2. Mientras existieran valores no nulos en los vectores de energía originales:
  - Se seleccionó aleatoriamente uno de los nueve canales. (Esta selección aleatoria permite que si dos cantos pertenecientes a individuos distintos se multiplexan en tiempo, se guarde la información de ambos en el *dataset* final.)
  - Se halló la posición  $j^*$  de la energía máxima en ese canal.
  - Se guardó el valor de dicha energía en la posición  $j^*$  del *array* de resultados para ese micrófono.
  - Se estableció a cero el valor de la energía en la posición  $j^*$  de todos los vectores, eliminando así la contribución de ese canto para las demás grabaciones.
3. El proceso se repitió hasta que todos los vectores de energía quedaron vacíos.
4. Finalmente, se almacenaron los *arrays* resultantes como secuencias de cantos por micrófono.

Este procedimiento, ilustrado en el pseudocódigo del Algoritmo 1, garantizó que cada instante de canto fuera asignado al micrófono cuya energía fuera dominante. La complejidad temporal resultó  $O(n^2)$ , con  $n$  igual al número de muestras en el eje temporal del Mel-espectrograma.

**Algorithm 1** Colin Sequence by Energy**Require:** Temporal energies  $\{E^{(k)}\}_{k=1}^9$ , length  $N$ **Ensure:** Calls sequences  $\{S^{(k)}\}_{k=1}^9$ 


---

```

1: Initialize  $S^{(k)} \leftarrow [0, \dots, 0]$  for  $k = 1, \dots, 9$ 
2: while  $\exists k, j \mid E_j^{(k)} > 0$  do
3:   Select randomly  $k \in \{1, \dots, 9\}$ 
4:    $j^* \leftarrow \arg \max_j E_j^{(k)}$ 
5:    $S_{j^*}^{(k)} \leftarrow E_{j^*}^{(k)}$  ▷ Save maximum value to file
6:   for each  $k' = 1, \dots, 9$  do ▷ Erase value from all arrays
7:      $E_{j^*}^{(k')} \leftarrow 0$ 
8:   end for
9: end while
10: return  $\{S^{(k)}\}$ 

```

---

**2.4.2. Algoritmo basado en agrupamiento espectro-temporal**

El segundo método propuesto partió también de los Mel-espectrogramas preprocesados y filtrados, pero incorporó explícitamente la estructura temporal y espectral de los cantos “CO” y “LIN” de *E. eileenae*. La idea central consistió en extraer todos los máximos locales del espectrograma (cada uno identificado por un par  $(t, f)$ ) y luego agruparlos en parejas correspondientes a los dos componentes del canto (frecuencia baja para “CO” y alta para “LIN”) de acuerdo con ventanas temporales aprendidas a partir de histogramas de datos manuales. A las parejas  $(t_{\text{CO}}, f_{\text{CO}})$  y  $(t_{\text{LIN}}, f_{\text{LIN}})$  obtenidas se les aplicó un algoritmo aglomerativo de clustering, usando como métrica la suma de cuadrados de la diferencia de frecuencias, de modo que aquellas parejas cuyos valores espectrales y temporales se encontraran más cercanos quedaran agrupadas. Cada grupo resultante definió una hipótesis de secuencia de canto, a la que se asignó la potencia media más alta como criterio de selección de la rana más cercana. Finalmente, los canales en los que el número de máximos agrupados no alcanzó un umbral mínimo fueron descartados, asumiendo la ausencia de un Colín próximo. A continuación se describen los pasos seguidos, también ilustrados en el pseudocódigo del Algoritmo 2.

1. Se calculó el Mel-espectrograma de cada archivo.
2. Se localizaron todos los máximos locales del Mel-espectrograma, cada



## 2.5. MODELADO DE INTERACCIONES CON EL MODELO DE ISING33

uno especificado por un par  $(t, f)$ .

3. Se filtraron aquellos máximos cuyas frecuencias y tiempos coincidían con los rangos propios de las notas “CO” y “LIN”, según los histogramas derivados de datos reales.
4. A las parejas  $(t_{CO}, f_{CO})$  y  $(t_{LIN}, f_{LIN})$  resultantes se les aplicó clustering aglomerativo, con distancia basada en la diferencia al cuadrado de frecuencias.
5. Se seleccionó la secuencia cuyo grupo presentó la potencia media de canto más alta, asignándola al micrófono correspondiente.
6. Se descartaron los canales con un número de máximos inferior al umbral mínimo, declarándolos sin detección de Colín.

---

**Algorithm 2** Spectral-Time Clustering Algorithm for Frog Call Detection

---

**Require:** Audio files `audio_files`, minimum cluster size threshold `min_threshold`

- 1: **for** each `audio_file` in `audio_files` **do**
- 2:   Compute Mel-spectrogram: `S = melspectrogram(audio_file)`
- 3:   Identify all local maxima in `S` as pairs  $(t, f)$
- 4:   Filter maxima to keep only those within CO and LIN frequency and time ranges based on annotated histograms
- 5:   Group pairs  $(t_{CO}, f_{CO})$  and  $(t_{LIN}, f_{LIN})$  into sequences
- 6:   Apply agglomerative clustering to these sequences using squared frequency difference as distance metric
- 7:   Compute mean power for each cluster
- 8:   **if** maximum cluster size  $< \text{min\_threshold}$  **then**
- 9:     Discard this audio file (no frog detected)
- 10:   **else**
- 11:     Assign the cluster with the highest mean power to this microphone
- 12:   **end if**
- 13: **end for**

---

## 2.5. Modelado de interacciones con el modelo de Ising

El análisis de las dinámicas colectivas de los Colines requiere un marco formal capaz de capturar las dependencias entre los individuos a partir de

observaciones de su comportamiento acústico. En esta sección se describe el uso del modelo de Ising como herramienta para representar y cuantificar las interacciones entre machos vocalizantes, siguiendo un enfoque basado en inferencia estadística y física computacional.

### 2.5.1. Fundamentos del modelo de Ising

El modelo de Ising fue concebido por Wilhelm Lenz e introducido por Ernst Ising en 1925 como una simplificación del comportamiento colectivo de espines en materiales ferromagnéticos [23]. En su formulación clásica, cada componente del sistema está representado por una variable binaria  $\sigma_i \in \{-1, +1\}$ , que puede interpretarse como un estado de activación (por ejemplo, presencia o ausencia de vocalización). Las interacciones entre pares de componentes se codifican mediante parámetros  $J_{ij}$ , que determinan si las unidades tienden a sincronizarse o a inhibirse mutuamente, mientras que un parámetro  $h_i$  refleja la predisposición individual (o sesgo) del componente  $i$  a estar activo.

El conjunto de estados del sistema está dado por el vector binario  $\sigma = (\sigma_1, \dots, \sigma_N)$ , y la probabilidad de observar una configuración determinada se modela según la distribución de Boltzmann:

$$P(\sigma) = \frac{1}{Z} \exp \left( \sum_{i < j} J_{ij} \sigma_i \sigma_j + \sum_i h_i \sigma_i \right),$$

donde  $Z$  es la función de partición que normaliza la distribución:

$$Z = \sum_{\sigma} \exp \left( \sum_{i < j} J_{ij} \sigma_i \sigma_j + \sum_i h_i \sigma_i \right).$$

Esta formulación permite representar redes de interacciones con estructuras arbitrarias, haciendo del modelo de Ising una herramienta versátil que ha sido empleada exitosamente en áreas como neurociencia, genética, economía y análisis de redes sociales [10].

### 2.5.2. Formulación del sistema acústico como red de espines

Para modelar los coros de *E. eileenae*, se asumió que cada uno de los nueve individuos registrados corresponde a un nodo del sistema, y su comportamiento acústico en un instante dado queda representado por un espín  $\sigma_i \in \{-1, +1\}$ , donde  $+1$  indica que el individuo está emitiendo un canto y  $-1$  que permanece en silencio.

Las configuraciones  $\sigma$  fueron extraídas a partir de las secuencias temporales de detección previamente construidas, dividiendo el tiempo en intervalos discretos (frames) de longitud fija. Cada configuración refleja el estado coral del sistema en un instante, y el conjunto de configuraciones observado se utilizó como base para inferir los parámetros  $\{J_{ij}\}$  y  $\{h_i\}$  que describen la estructura de interacción y la propensión individual a cantar.

Aunque el sistema de cantos no necesariamente satisface los criterios de equilibrio termodinámico, se adoptó esta suposición como primera aproximación, lo cual permitió el uso de herramientas formales del marco de física estadística [49]. Esta simplificación habilita una descripción cuantitativa de las correlaciones observadas, útil para identificar influencias directas entre individuos del coro.

### 2.5.3. Inferencia por máxima verosimilitud y descenso por gradiente

La tarea de inferir los parámetros del modelo a partir de observaciones consiste en encontrar los valores de  $J_{ij}$  y  $h_i$  que maximizan la verosimilitud del conjunto de configuraciones observadas:

$$\mathcal{L}(J, h) = \sum_{m=1}^M \log P(\sigma^{(m)}),$$

donde  $\{\sigma^{(m)}\}_{m=1}^M$  son las configuraciones registradas.

El gradiente de esta función de verosimilitud con respecto a los parámetros tiene la forma:

$$\frac{\partial \mathcal{L}}{\partial J_{ij}} = \langle \sigma_i \sigma_j \rangle_{\text{datos}} - \langle \sigma_i \sigma_j \rangle_{\text{modelo}},$$

$$\frac{\partial \mathcal{L}}{\partial h_i} = \langle \sigma_i \rangle_{\text{datos}} - \langle \sigma_i \rangle_{\text{modelo}},$$

donde los términos  $\langle \cdot \rangle_{\text{datos}}$  se obtienen como promedios sobre las configuraciones empíricas, y los términos  $\langle \cdot \rangle_{\text{modelo}}$  se calculan a partir de la distribución del modelo.

En este trabajo, gracias al tamaño reducido del sistema (9 nodos,  $2^9 = 512$  configuraciones), fue posible calcular de forma exacta la función de partición y las expectativas modeladas, evitando el uso de muestreo estocástico como Monte Carlo [30]. Esto permitió implementar un algoritmo de descenso por gradiente exacto, complementado con regularización  $L_2$  sobre los pesos  $J_{ij}$  para evitar sobreajuste.

## 2.5. MODELADO DE INTERACCIONES CON EL MODELO DE ISING36

La energía asociada a una configuración se definió como:

$$E(\sigma) = - \sum_{i < j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i,$$

lo cual permitió expresar su probabilidad como:

$$P(\sigma) = \frac{1}{Z} \exp(-E(\sigma)).$$

El algoritmo iterativo actualizó los parámetros siguiendo las reglas:

$$h_i \leftarrow h_i + \eta (\langle \sigma_i \rangle_{\text{datos}} - \langle \sigma_i \rangle_{\text{modelo}}), \quad (2.2)$$

$$J_{ij} \leftarrow J_{ij} + \eta (\langle \sigma_i \sigma_j \rangle_{\text{datos}} - \langle \sigma_i \sigma_j \rangle_{\text{modelo}}) - \eta \lambda J_{ij}, \quad (2.3)$$

donde  $\eta$  es la tasa de aprendizaje y  $\lambda$  el coeficiente de regularización. Además para el cálculo de las correlaciones empíricas se utilizó un  $\Delta t = 0$ , es decir, se tuvo en cuenta que cada estado era el valor de canto o no canto (+1 o -1) en un instante de tiempo.

---

**Algorithm 3** Gradient Descent for Ising Model Parameter Inference

---

**Require:** Observed configurations  $\{\sigma^{(1)}, \dots, \sigma^{(M)}\}$ , learning rate  $\eta$ , number of iterations  $T$ , regularization parameter  $\lambda$

- 1: Initialize  $h_i \leftarrow 0, J_{ij} \leftarrow 0$  for all  $i, j$
- 2: **for**  $t \leftarrow 1$  to  $T$  **do**
- 3:   Compute empirical averages:

$$\langle \sigma_i \rangle_{\text{data}}, \quad \langle \sigma_i \sigma_j \rangle_{\text{data}}$$

- 4:   Enumerate all configurations  $\sigma \in \{-1, +1\}^9$
- 5:   **for** each configuration  $\sigma$  **do**
- 6:     Compute energy:

$$E(\sigma) = - \sum_{i < j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i$$

- 7:     Compute unnormalized probability:

$$\tilde{P}(\sigma) = \exp(-E(\sigma))$$

- 8:   **end for**
- 9:   Normalize to obtain  $P(\sigma)$
- 10:   Compute model expectations:

$$\langle \sigma_i \rangle_{\text{model}}, \quad \langle \sigma_i \sigma_j \rangle_{\text{model}}$$

- 11:   Update parameters:

$$h_i \leftarrow h_i + \eta (\langle \sigma_i \rangle_{\text{data}} - \langle \sigma_i \rangle_{\text{model}})$$

$$J_{ij} \leftarrow J_{ij} + \eta (\langle \sigma_i \sigma_j \rangle_{\text{data}} - \langle \sigma_i \sigma_j \rangle_{\text{model}}) - \eta \lambda J_{ij}$$

- 12: **end for**
- 

Este procedimiento (descrito en el pseudocódigo del Algoritmo 3) garantizó la convergencia a un conjunto de parámetros que maximizan la verosimilitud del modelo, y proporciona una primera reconstrucción de la red de interacciones acústicas dentro del coro.

## Capítulo 3

# Resultados e Interpretación

Antes de presentar los resultados, es importante recordar que el flujo completo de procesamiento (desde la sincronización y el filtrado de ruido hasta la detección y asignación heurística de los cantos) se diseñó con el fin de reconstruir con precisión las series temporales de vocalización de cada individuo. En este capítulo se evalúa, en primer lugar, la consistencia y precisión logradas por los algoritmos de detección y asignación micrófono-Colín, comparando su rendimiento con mediciones manuales de referencia. A continuación, se efectúa una comparación entre las diferentes estrategias heurísticas desarrolladas, con el objetivo de identificar fortalezas y limitaciones de cada enfoque. Posteriormente, se analiza la red de interacciones inferidas (los parámetros  $\{J_{ij}\}$ ) para caracterizar la estructura acústica del coro. Finalmente, se discute la idoneidad del modelo de Ising como representación estadística del sistema, valorando su capacidad para reproducir los patrones de cantos observados.

### 3.1. Resultados y evaluación de los algoritmos propuestos

Al aplicar el algoritmo descrito en la Sección 2.4.1 sobre el *dataset* preprocesado, se obtuvo, para cada hora, una secuencia temporal de los cantos de cada Colín. Un fragmento representativo de diez segundos de los archivos correspondientes a 20231021\_190000 se muestra en la Figura 3.1, donde puede apreciarse cómo el método distingue de forma coherente los eventos de “CO” y “LIN” en cada canal.

En el intervalo analizado, las detecciones mantuvieron una periodicidad

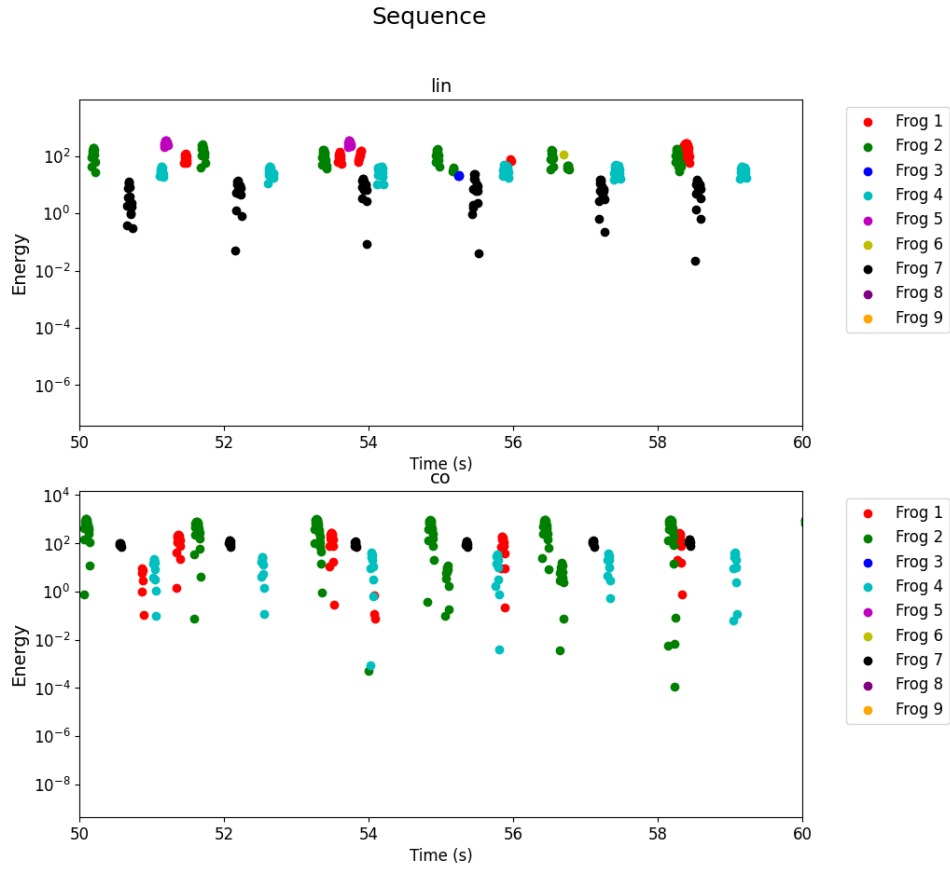


Figura 3.1: Fragmento de Secuencia de Cantos de los Colines en los archivos de 20231021\_190000. Arriba la señales “LIN”, debajo los “CO”. Cada color representa una rana o micrófono diferente.

dad aproximada y una consistencia energética que parecen indicar la correcta detección y asignación de los cantos, condición imprescindible para avanzar hacia el estudio de las interacciones acústicas.

### 3.1.1. Consistencia

Para evaluar la consistencia del algoritmo, se realizaron diez ejecuciones independientes sobre el mismo conjunto de datos. En cada corrida se generaron nueve vectores (uno por micrófono-Colín), cada uno con los va-

lores de energía de los cantos asignados a ese micrófono, ubicados en los índices correspondientes a los instantes de tiempo en que fueron detectados los cantos. A fin de cuantificar la estabilidad de los resultados, se construyó una gran matriz de correlación cruzada entre todos los vectores provenientes de las diferentes ejecuciones.

La organización de la matriz siguió un orden bloque por bloque: primero se agruparon los vectores del micrófono “a” en cada una de las diez corridas, luego los del micrófono “b”, y así sucesivamente. De este modo, una alta correlación a lo largo de la diagonal principal indicaría que, para un mismo micrófono, las detecciones fueron reproducibles en cada ejecución.

La Figura 3.2 presenta el mapa de calor de esa matriz. Se resaltaron en rojo las diez correlaciones más elevadas de cada fila. La evidente concentración de valores altos en la diagonal principal confirma que los vectores de detección correspondientes a un mismo micrófono exhibieron correlaciones significativamente mayores entre sí que con los vectores de otros canales. Este comportamiento empírico demuestra que el algoritmo produjo resultados coherentes y reproducibles.

### 3.1.2. Comparación con detecciones manuales

Para evaluar de forma cualitativa el desempeño de ambos algoritmos, se compararon sus detecciones con un etiquetado manual de cantos “CO” y “LIN” realizado sobre el conjunto de grabaciones de la sesión 20231021\_190000. Dado que el etiquetado fue efectuado por personal no experto, esta comparación se abordó de manera visual y comparativa, sin aspirar a la rigurosidad de un protocolo validado por especialistas.

En la Figura 3.3 se presenta la matriz de confusión del algoritmo basado en energías relativas (izq.) y la correspondiente al algoritmo de *clustering* espectro-temporal (der.). Para esto se tomaron en cuenta solamente las grabaciones donde se detectaron manualmente una gran cantidad de cantos. Se consideró que un algoritmo había detectado correctamente un canto cuando coincidía temporalmente (con tolerancia de 1 segundo) y en micrófono con los datos etiquetados. Se observa que el método de *clustering* obtuvo un mayor número de verdaderos positivos y menos falsos positivos que el de energías relativas. Ello se explica porque, cuando no todas las ranas estaban activas, el algoritmo de energías relativas tendía a asignar cantos a cada micrófono (incluso en aquellos donde no se encontraba cantando ningún colín), elevando así sus falsos positivos y negativos. Por el contrario, el *clustering* descartó los canales con actividad insuficiente, resultando más



### 3.1. RESULTADOS Y EVALUACIÓN DE LOS ALGORITMOS PROPUESTOS 41

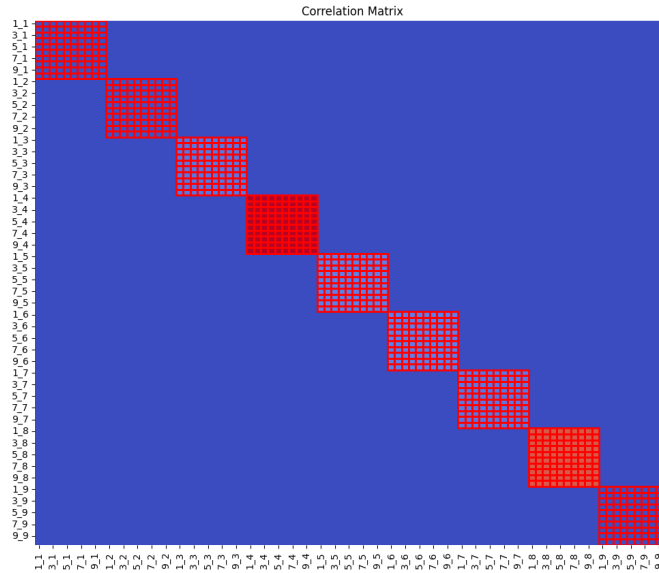


Figura 3.2: Matriz de correlación entre ejecuciones independientes del algoritmo sobre los datos de 20231021\_190000. Los bloques diagonales de alta correlación (marcados en rojo) muestran la consistencia de las detecciones para cada micrófono.

robusto en entornos de baja densidad acústica.

No se contabilizaron los verdaderos negativos (no canto) dado que dichos valores no forman parte de la salida de los algoritmos.

Adicionalmente, se compararon los  $\Delta t$  (tiempo entre “CO” y su “LIN” correspondiente) obtenidos manualmente y por cada algoritmo. En el caso del método de energías, al ejecutarse por separado en los “CO” y los “LIN”, algunos “CO” o “LIN” quedaron desemparejados y fueron filtrados mediante un umbral extraído del etiquetado manual. Tras este emparejamiento, el número de llamados considerados por el algoritmo de energías relativas bajó de 7 290 “LIN” y 9 863 “CO” a 5 467 pares, cifra cercana a los 5 964 pares hallados por *clustering*.

Los *boxplots* y *violinplots* de la Figura 3.4 muestran que la distribución de  $|\Delta t|$  es muy similar entre el etiquetado manual, el algoritmo de energías y el de *clustering*. Asimismo, los histogramas comparativos (Figura 3.5) revelan alineamientos próximos de ambas técnicas con el patrón de referencia, a pesar de la posible imprecisión en los tags manuales.

En conjunto, estos resultados sugieren que, comparándose con un eti-

### 3.1. RESULTADOS Y EVALUACIÓN DE LOS ALGORITMOS PROPUESTOS 42

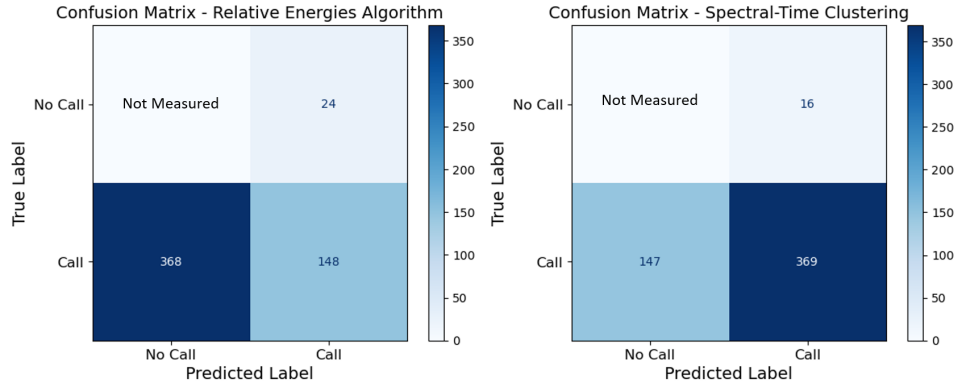


Figura 3.3: Matriz de confusión del algoritmo de energías relativas vs. detecciones manuales (izq.). Matriz de confusión del algoritmo de *clustering* espectro-temporal vs. detecciones manuales (der.). Solo se consideraron los cantos detectados en las grabaciones con mayor actividad.

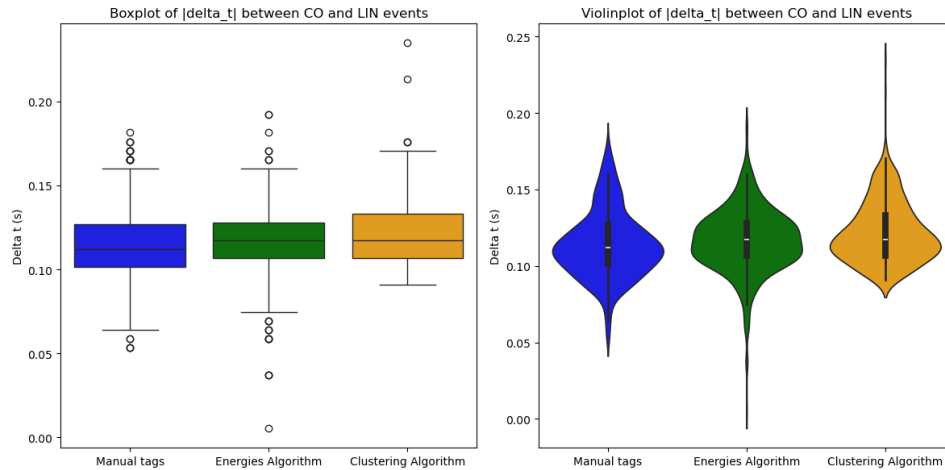


Figura 3.4: *Boxplot* y *violinplot* de  $|\Delta t|$  entre “CO” y “LIN” para detecciones manuales, algoritmo de energías relativas y *clustering*.

quetado manual inexperto, el algoritmo basado en *clustering* muestra mejor rendimiento en cuanto a capturar la dinámica temporal del canto, pues ofrece mayor fiabilidad en escenarios de baja actividad. Para una valida-

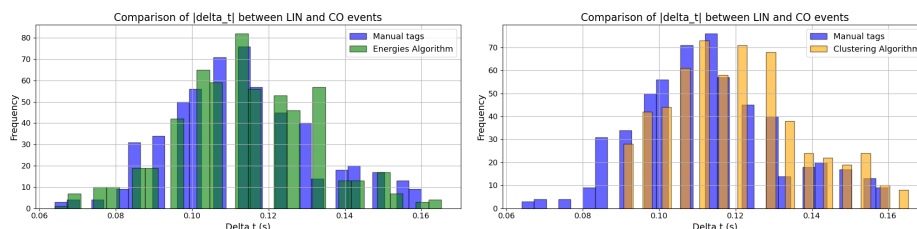


Figura 3.5: Histogramas de  $\Delta t$ : (izq.) manual vs. energías relativas; (der.) manual vs. *clustering*. Los colores representan el origen de los datos (etiquetado manual, algoritmo de energías relativas o algoritmo de *clustering*).

ción estadística más rigurosa sería necesario un etiquetado experto que minimice el ruido de referencia.

### 3.2. Comparación entre algoritmos

Con el objetivo de validar la robustez de los métodos propuestos para la detección y asignación de cantos, se compararon los resultados obtenidos mediante ambos algoritmos heurísticos presentados en el Capítulo anterior. La Figura 3.6 muestra un gráfico de dispersión que compara los tiempos de detección de cada canto individual en los archivos de 20231021\_190000, con cada punto representando un canto detectado por ambos algoritmos y coloreado según el individuo correspondiente. La línea negra discontinua representa la identidad  $y = x$ , indicando perfecta coincidencia entre ambos métodos. Por simplicidad, el algoritmo descrito en la Sección 2.4.1 (basado en energías relativas) se denominó **Algoritmo A** y el de la Sección 2.4.2 (*clustering* de espectro-tiempo), **Algoritmo B**.

Como puede observarse, la mayoría de los puntos se alinean con alta precisión sobre la recta de identidad, lo que indica un alto grado de concordancia entre ambos algoritmos en cuanto a los instantes de detección. Este resultado sugiere que, a pesar de basarse en supuestos distintos, ambos métodos producen secuencias temporales coherentes, lo cual refuerza la validez de los procedimientos desarrollados.

Para cuantificar con mayor precisión esta similitud, se realizó un ajuste lineal entre los tiempos de detección generados por ambos algoritmos para cada canal (o micrófono) por separado. Se calculó el coeficiente de determinación  $R^2$ , métrica comúnmente empleada para evaluar la calidad del ajuste de modelos lineales [28]. Un valor cercano a  $R^2 = 1$  reflejan una alta

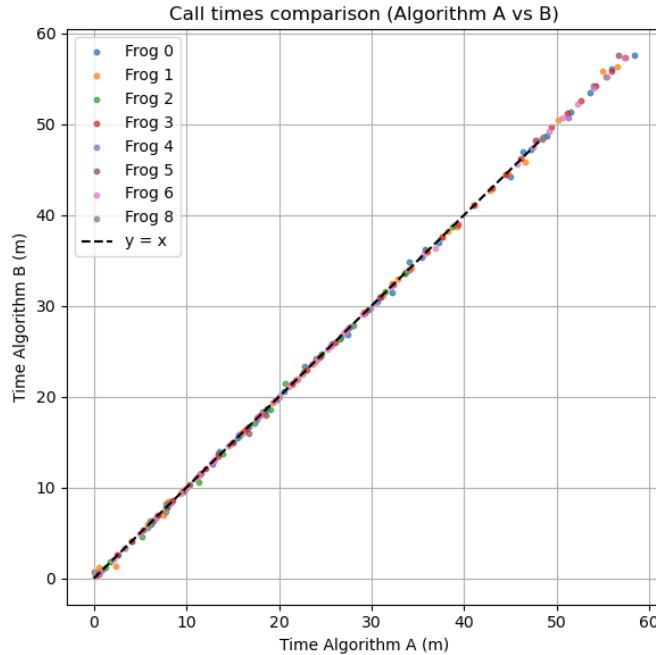


Figura 3.6: Comparación entre los tiempos de detección de cantos por los algoritmos A (energías relativas) y B (comportamiento de la especie) sobre los datos de 20231021\_190000.

correlación.

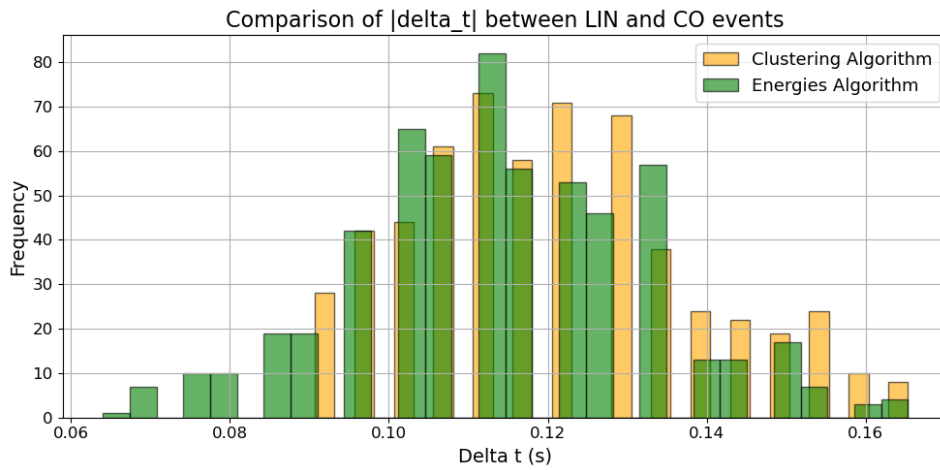
La Tabla 3.1 resume los valores obtenidos de  $R^2$  para los nueve canales analizados. En ocho de los casos se obtuvo un valor de  $R^2 = 1,00$ , indicando un solapamiento perfecto entre los resultados. Solo uno de los canales presentó un valor ligeramente inferior ( $R^2 = 0,99$ ), lo cual sigue siendo una indicación de coincidencia prácticamente total.

Estos resultados corroboran que los algoritmos propuestos no solo son internamente consistentes, sino que ofrecen una detección altamente reproducible y concordante, independientemente del criterio heurístico considerado. Esta coincidencia refuerza la fiabilidad de las secuencias de canto extraídas, base fundamental para el análisis posterior de las interacciones acústicas en el sistema.

Adicionalmente, se analizaron las discrepancias en la detección de eventos “CO-LIN” entre ambos métodos. La Figura 3.7 presenta el histograma comparativo de los  $\Delta t$  (diferencia temporal entre “CO” y “LIN”) detectados por el Algoritmo A y el Algoritmo B.

**Cuadro 3.1** Coeficientes de determinación  $R^2$  del ajuste lineal entre algoritmos A y B por micrófono.

Micrófono	$R^2$
20231021_190000a	1.00
20231021_190000b	1.00
20231021_190000c	1.00
20231021_190000d	1.00
20231021_190000e	0.999
20231021_190000f	1.00
20231021_190000g	1.00
20231021_190000h	1.00
20231021_190000i	1.00

Figura 3.7: Histogramas comparativos de  $\Delta t$  entre “CO” y “LIN” para el Algoritmo A (energías temporales) y el Algoritmo B (*clustering* de espectro-tiempo).

Por último, se contabilizó cuántos cantos fueron detectados exclusivamente por cada algoritmo y cuántos coincidieron temporalmente (dentro de  $\pm 1$  s) en el mismo micrófono. Los resultados fueron:

- Solo detectados por Algoritmo A: 2429 cantos
- Solo detectados por Algoritmo B: 3000 cantos

- Detectados por ambos algoritmos: 3038 cantos

Estas cifras indican que el Algoritmo B (*clustering*) identificó un mayor número de cantos no captados por el Algoritmo A, mientras que ambos detectaron conjuntamente la mayoría de eventos (3038). En particular, el Algoritmo A pasó por alto una parte significativa de cantos (2429 únicos de B frente a 3000 únicos de A), lo que refuerza la hipótesis de que su estrategia de energías relativas resulta menos sensible cuando no todos los individuos emiten simultáneamente. Por el contrario, el *clustering* demostró una mayor cobertura de detección sin sacrificar la coincidencia temporal, confirmando su idoneidad en situaciones de actividad coral variable.

En conjunto, esta comparación demuestra que los dos algoritmos proporcionan resultados muy similares en cuanto a tiempo de detección ( $R^2$  casi unidad), pero difieren en sensibilidad: especialmente cuando no todos los Colines estaban activos simultáneamente, el método basado en *clustering* (Algoritmo B) mostró mejor rendimiento general.

### 3.3. Análisis de las interacciones inferidas $\{J_{ij}\}$

La red de interacciones acústicas entre individuos fue inferida mediante el modelo de Ising descrito en el Capítulo anterior, utilizando configuraciones binarizadas de cantos. Este modelo permitió cuantificar la fuerza de acoplamiento entre pares de individuos, interpretada como una medida de influencia mutua en la emisión de vocalizaciones.

En la Figura 3.8 se muestra la matriz simétrica  $J_{ij}$  obtenida para los datos de 20232110\_190000. En ella, cada entrada representa la magnitud y dirección de la interacción inferida entre los individuos  $i$  y  $j$ . Se observa una estructura dispersa, con varios acoplamientos claramente distintos de cero, lo que sugiere la existencia de vínculos acústicos no triviales entre ciertos miembros del coro. Las intensidades positivas indican una tendencia a la co-emisión simultánea, mientras que los valores negativos reflejan inhibición recíproca.

Con el fin de facilitar la interpretación estructural de estas relaciones, se construyó un grafo no dirigido donde cada nodo representa a un individuo (o micrófono) y las aristas indican la existencia de interacciones significativas. En particular, se utilizó una línea continua para aquellas interacciones cuya magnitud supera el umbral de 0.5, y una línea discontinua para aquellas en el rango entre 0.3 y 0.5. Las conexiones cuyas magnitud resultaron inferiores a este último valor se consideraron insignificantes y fueron omi-

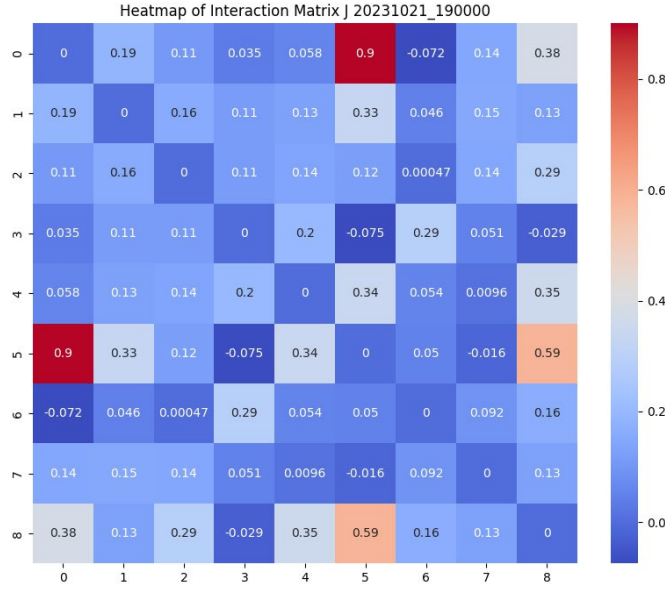


Figura 3.8: Matriz de interacciones  $J_{ij}$  inferida para el 20231021\_190000.

tidas del grafo para mejorar la claridad visual. El resultado se presenta en la Figura 3.9.

Como análisis exploratorio adicional, se aplicó el mismo procedimiento de inferencia sobre tres intervalos horarios consecutivos de la misma noche (20231021\_180000, 20231021\_190000, 20231021\_200000). La Figura 3.10 muestra la evolución de las matrices  $J_{ij}$  a lo largo de estas tres horas. Se observa que varias de las interacciones más fuertes (especialmente aquellas con valores superiores a 0.5) se mantienen estables en el tiempo. Un ejemplo destacado es la interacción entre los individuos indexados como 0 y 5, que presenta una magnitud elevada y sostenida a lo largo del intervalo temporal analizado.

Este patrón de persistencia fue también evidente en la representación gráfica de las redes inferidas, ilustradas en la Figura 3.11. La presencia continua de ciertas aristas a lo largo del tiempo sugiere la existencia de relaciones acústicas recurrentes y posiblemente funcionales entre algunos individuos. Dichos vínculos podrían reflejar mecanismos de coordinación o competencia territorial que merecen ser explorados en mayor profundidad en estudios futuros.

Bajo el supuesto de que el modelo de Ising logra cantificar correcta-

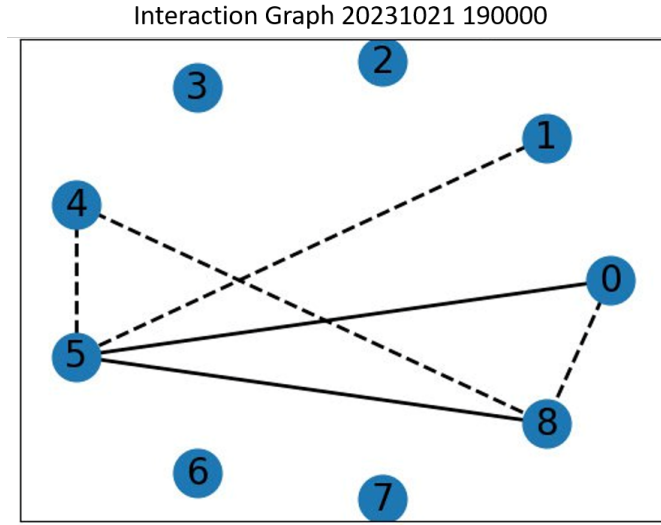


Figura 3.9: Grafo de interacciones correspondiente al 20231021\_190000.

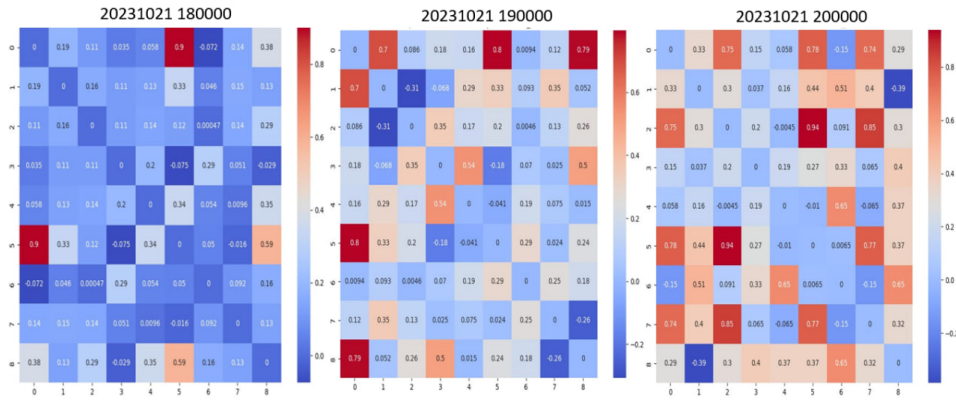


Figura 3.10: Evolución de la matriz  $J_{ij}$  durante tres horas consecutivas (20231021\_180000, 20231021\_190000, 20231021\_200000).

mente las interacciones del sistema, estos hallazgos refuerzan la hipótesis de que los coros de *E. eileenae* exhiben una estructura interna no aleatoria.



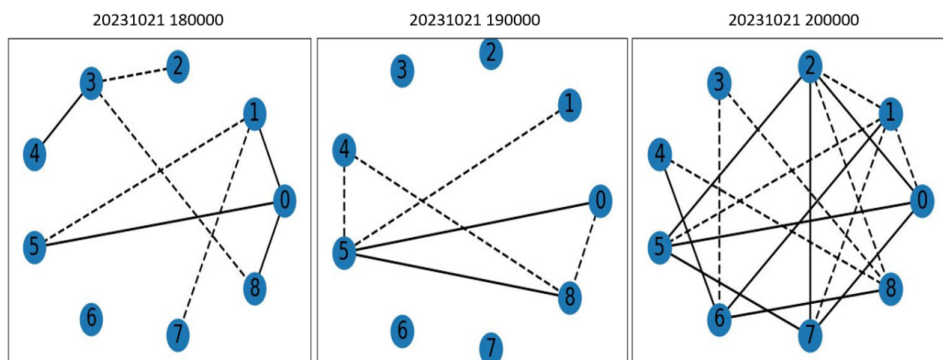


Figura 3.11: Evolución temporal de los grafos de interacciones inferidas para tres horas consecutivas (20231021\_180000, 20231021\_190000, 20231021\_200000).

### 3.4. Análisis de idoneidad del modelo de Ising

Para evaluar la capacidad predictiva del modelo de Ising, se compararon sus estimaciones de la tasa de aparición de patrones de canto con las obtenidas bajo un modelo independiente en el que se asumen que cada Colín actúa sin influencia de los otros individuos [35].

En la Figura 3.12 se presentan, en escala logarítmica, los patrones observados frente a los patrones predichos por ambos modelos para dos ventanas horarias distintas: 20231021\_190000 y 20231021\_040000. Cada punto corresponde a la frecuencia de un patrón particular de activación de espines (combinación de individuos cantando simultáneamente). La línea punteada  $y = x$  indica coincidencia perfecta entre observación y predicción.

Contrariamente a lo esperado, el modelo independiente mostró una correspondencia más estrecha con las tasas observadas que el modelo de Ising, cuya predicción se desvió considerablemente en varios patrones de baja frecuencia. Este resultado sugiere al menos tres posibles interpretaciones: (1) las interacciones acústicas entre individuos podrían ser efectivamente débiles o nulas durante las ventanas analizadas, (2) el modelo de Ising, en su forma estacionaria y de segundo orden, podría no capturar adecuadamente la dinámica temporal o no lineal de los coros, o (3) la escasez de datos para patrones raros impide una estimación fiable de los parámetros de interacción. Esta última parece la más probable, dado que en la ejecución del descenso por gradiente se escogió un  $\Delta t = 0$  para cuantificar

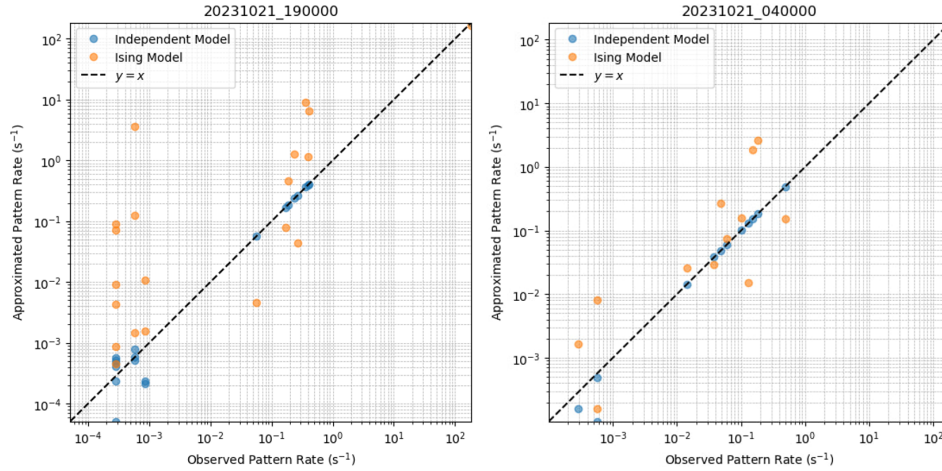


Figura 3.12: Tasa de aparición de patrones de canto observada versus predicha por el modelo independiente (azul) y el modelo de Ising (naranja) para las grabaciones de 20231021\_190000 (izq.) y 20231021\_040000 (der.).

las correlaciones entre los espines (los Colines) a través de los estados del sistema. Esto puede estar arrojando resultados engañosos debido a que es mucho más probable un estado de -1 que de +1 (es mucho más tiempo en el que los Colines no vocalizan que el tiempo en que lo hacen).

En conjunto, estos hallazgos indican que, para el conjunto de grabaciones y la metodología empleada, el supuesto de independencia entre Colines resulta tan (o más) efectivo que el modelo de Ising para describir la distribución de patrones de canto. No obstante, la presencia de acoplamientos significativos observados en Sección 3.3 sugiere que podría ser necesaria una extensión del modelo (por ejemplo, incorporando términos de orden superior, efectos no estacionarios o dependencias temporales explícitas, así como explorar los valores del umbral de tiempo en el que se calculan las correlaciones) para capturar plenamente la complejidad del comportamiento grupal de *E. eileenae*.

# Conclusiones

El presente estudio desarrolló y validó un flujo computacional automatizado para el análisis de coros de *Eleutherodactylus eileenae* a partir de grabaciones bioacústicas. Los objetivos específicos se abordaron mediante tres etapas principales: eliminación de ruido y sincronización de señales, detección y asignación de cantos, e inferencia de interacciones acústicas. La sincronización de las nueve pistas de audio se logró mediante correlación cruzada sobre los histogramas de tiempos calculados a partir de los puntos de máximos locales en los Mel-espectrogramas, propiciando la correcta aplicación de los métodos de detección. Los algoritmos heurísticos propuestos para detectar y asignar cantos demostraron alta consistencia interna, con correlaciones superiores elevadas entre ejecuciones independientes. El método basado en *clustering* espectro-temporal mostró mayor sensibilidad en escenarios de baja actividad, superando al enfoque de energías relativas.

La inferencia de interacciones mediante el modelo de Ising reveló acoplamientos significativos entre individuos, evidenciando una estructura de red con tendencias tanto a la co-emisión como a la inhibición. Sin embargo, la comparación con un modelo independiente sugirió que las interacciones podrían ser débiles, que los patrones se manifiestan con poca frecuencia en los datos, lo que los hace difíciles de predecir, o que el modelo de Ising, en su formulación estándar, no captura plenamente la dinámica no lineal del sistema. Esto subraya la necesidad de incorporar términos de orden superior o dinámicas temporales explícitas en futuras aproximaciones. Además, la validación manual limitada por la ausencia de un *ground truth* experto resaltó la importancia de mejorar los protocolos de etiquetado para reducir sesgos en la evaluación de algoritmos.

El presente trabajo representa un avance significativo en el estudio de *Eleutherodactylus eileenae* en Cuba, al introducir un flujo completamente automatizado para el procesamiento de las grabaciones de campo, capaz de identificar distintos individuos dentro de una misma especie de anuros, en contraste con la mayoría de estudios previos que se han centrado en

distinguir especies distintas a partir de sus vocalizaciones. Además, se propuso un enfoque novedoso para modelar matemáticamente las interacciones acústicas del coro mediante un sistema de espines. Con ello se facilita el análisis reproducible de grandes volúmenes de datos bioacústicos y se sienta una base cuantitativa para investigar la dinámica colectiva de los machos cantores.

# Recomendaciones

A partir de los hallazgos y limitaciones identificadas, se proponen las siguientes recomendaciones para futuras investigaciones:

- **Facilitar la sincronización y validación:** Incluir pulsos acústicos artificiales (tonos de referencia) al inicio y final de cada grabación para facilitar la alineación temporal. Paralelamente, colaborar con bioacústicos expertos en *E. eileenae* para generar un *ground truth* robusto mediante etiquetado manual estandarizado, permitiendo métricas de evaluación más rigurosas.
- **Ampliar la diversidad de datos:** Replicar el estudio en múltiples temporadas reproductivas y localidades geográficas, incorporando variables ambientales (temperatura, humedad) para analizar su influencia en la estructura de los coros. Esto enriquecería la generalidad de los modelos propuestos.
- **Integrar información espacial:** Utilizar micrófonos direccionales o arreglos de sensores para triangular posiciones exactas de los individuos, combinando datos acústicos con coordenadas georreferenciadas. Esta información permitiría modelar interacciones en función de distancias físicas.

Estas mejoras metodológicas, junto con la adopción de marcos teóricos más flexibles, potenciarían el estudio cuantitativo de sistemas bioacústicos complejos, tanto en anuros como en otras especies gregarias.

# Referencias Bibliográficas

- [1] Miguel A Acevedo, Carlos J Corrada-Bravo, Héctor Corrada-Bravo, Luis J Villanueva-Rivera, and T Mitchell Aide. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4):206–214, 2009. (Citado en las páginas 11 y 17).
- [2] Roberto Alonso, Ariel Rodríguez-Gómez, and Alberto R Estrada. Patrones de actividad acústica y trófica de machos cantores de *eleutherodactylus eileenae* (anura: Leptodactylidae). *Revista española de herpetología*, 15(2001):45–52, 2001. (Citado en las páginas 10, 12, 20 y 23).
- [3] Roberto Alonso Bosch and Ariel Rodríguez Gómez. Llamadas del repertorio vocal de *eleutherodactylus eileenae* (anura: Leptodactylidae) del occidente de cuba. *Cuadernos de Herpetología*, 14, 2001. (Citado en la página 26).
- [4] William Bialek. *Biophysics: searching for principles*. Princeton University Press, 2012. (Citado en la página 14).
- [5] Daniel T Blumstein, Daniel J Mennill, Patrick Clemins, Lewis Girod, Kung Yao, Gail Patricelli, Jill L Deppe, Alan H Krakauer, Christopher Clark, Kathryn A Cortopassi, et al. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48(3):758–767, 2011. (Citado en las páginas 10, 11 y 17).
- [6] Natalie T Boelman, Gregory P Asner, Patrick J Hart, and Roberta E Martin. Multi-trophic invasion resistance in hawaii: bioacoustics, field surveys, and airborne remote sensing. *Ecological Applications*, 17(8):2137–2144, 2007. (Citado en la página 20).

- [7] Sheryn Brodie, Slade Allen-Ankins, Michael Towsey, Paul Roe, and Lin Schwarzkopf. Automated species identification of frog choruses in environmental recordings using acoustic indices. *Ecological Indicators*, 119:106852, 2020. (Citado en la página 21).
- [8] Ryan Calsbeek, Francisco Javier Zamora-Camacho, and Laurel B Symes. Individual contributions to group chorus dynamics influence access to mating opportunities in wood frogs. *Ecology Letters*, 25(6):1401–1409, 2022.
- [9] Juan Sebastián Cañas, María Paula Toro-Gómez, Larissa Sayuri Moreira Sugai, Hernán Darío Benítez Restrepo, Jorge Rudas, Breyner Posso Bautista, Luís Felipe Toledo, Simone Dena, Adão Henrique Rosa Domingos, Franco Leandro de Souza, et al. A dataset for benchmarking neotropical anuran calls identification in passive acoustic monitoring. *Scientific Data*, 10(1):771, 2023. (Citado en la página 18).
- [10] H Chau Nguyen, Riccardo Zecchina, and Johannes Berg. Inverse statistical problems: from the inverse ising problem to data science. *Advances in Physics*, 66(3):197–261, 2017. (Citado en las páginas 22 y 34).
- [11] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965. (Citado en la página 18).
- [12] Cornell Lab of Ornithology. The birth of natural sound recording. All About Birds website, n.d. “Allen and Kellogg recorded a Song Sparrow, a House Wren and a Rose-breasted Grosbeak on May 18, 1929.”. (Citado en la página 17).
- [13] Cornell Lab of Ornithology. Early milestones (1920-1950). Macaulay Library website, n.d. “1929: Allen and Kellogg made the first recordings of wild birds in Ithaca.”. (Citado en la página 17).
- [14] Luciano da F Costa. Comparing cross correlation-based similarities. *arXiv preprint arXiv:2111.08513*, 2021. (Citado en las páginas 14 y 30).
- [15] George E Drewry and A Stanley Rand. Characteristics of an acoustic community: Puerto rican frogs of the genus *eleutherodactylus*. *Copeia*, pages 941–953, 1983. (Citado en la página 20).
- [16] AR Estrada. Nuevas localidades para *eleutherodactylus eileenae dunn* (amphibia: Anura: Leptodactylidae) de cuba. *Misc. Zool., Academia de Ciencias de Cuba*, 22(4), 1984. (Citado en la página 23).

- [17] AR Estrada. Herpetofauna de la cuenca banao-higuanojo, sancti spíritus, cuba. *Revista de la Academia Colombiana de Ciencias*, 19:353–360, 1994. (Citado en las páginas 10 y 23).
- [18] Almo Farina. Ecoacoustics: A quantitative approach to investigate the ecological role of environmental sounds. *Mathematics*, 7(1):21, 2018. (Citado en la página 21).
- [19] RA Fisher. The maximum–likelihood–method. *Messenger in Mathematics*, 41:155–160, 1912. (Citado en la página 15).
- [20] Hongxiao Gan, Jinglan Zhang, Michael Towsey, Anthony Truskinger, Debra Stark, Berndt J Van Rensburg, Yuefeng Li, and Paul Roe. A novel frog chorusing recognition method with acoustic indices and machine learning. *Future Generation Computer Systems*, 125:485–495, 2021. (Citado en la página 21).
- [21] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969. (Citado en la página 22).
- [22] Andrew P Hill, Peter Prince, Evelyn Piña Covarrubias, C Patrick Doncaster, Jake L Snaddon, and Alex Rogers. Audiomoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution*, 9(5):1199–1211, 2018. (Citado en la página 26).
- [23] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925. (Citado en las páginas 13, 14 y 34).
- [24] Daniel Jurafsky and James H Martin. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. (Citado en la página 31).
- [25] Lata Kalra. *Signal Perception in Complex Acoustic Environments: A Study on Treefrog Communication*. PhD thesis, University of Minnesota, 2024. (Citado en las páginas 18 y 19).
- [26] Paul Mermelstein. Distance measures for speech recognition, psychological and instrumental. *Pattern recognition and artificial intelligence*, 116:374–388, 1976. (Citado en la página 18).



- [27] Thierry Mora and William Bialek. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144:268–302, 2011. (Citado en las páginas 14 y 21).
- [28] Nico JD Nagelkerke et al. A note on a general definition of the coefficient of determination. *biometrika*, 78(3):691–692, 1991. (Citado en la página 43).
- [29] Kaiichiro Ota, Ikkyu Aihara, and Toshio Aoyagi. Interaction mechanisms quantified from dynamical features of frog choruses. *Royal Society open science*, 7(3):191693, 2020. (Citado en las páginas 18 y 20).
- [30] Daniel Peña Sánchez de Rivera. Deducción de distribuciones: el método de monte carlo. *Fundamentos de Estadística. Alianza Editorial, Madrid. pp*, pages 220–300, 2001. (Citado en la página 35).
- [31] Nadia Pieretti, Almo Farina, and Davide Morri. A new methodology to infer the singing activity of an avian community: The acoustic complexity index (aci). *Ecological indicators*, 11(3):868–873, 2011. (Citado en la página 20).
- [32] Alberto Porta and Luca Faes. Wiener–granger causality in network physiology with applications to cardiovascular control and neuroscience. *Proceedings of the IEEE*, 104(2):282–309, 2015. (Citado en la página 22).
- [33] A Reyes, M Curbelo, F Tejera, A Rivera, G Simon, O Ramos, MS Turner, and E Altshuler. Transmission of danger information past physical barriers by ants. *arXiv preprint arXiv:1904.03236*, 2019. (Citado en la página 14).
- [34] Ignacio Sánchez-Gendríz and Linilson Rodrigues Padovese. A methodology for analyzing biological choruses from long-term passive acoustic monitoring in natural areas. *Ecological Informatics*, 41:1–10, 2017.
- [35] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006. (Citado en las páginas 14, 21 y 49).
- [36] Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000. (Citado en la página 22).

- [37] Albert Schwartz. A new frog of the auriculatus group of the genus *eleutherodactylus* from western cuba. *Herpetologica*, 14(2):69–77, 1958. (Citado en las páginas 10 y 23).
- [38] John L Spiesberger and Kurt M Fristrup. Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography. *The american naturalist*, 135(1):107–153, 1990. (Citado en la página 19).
- [39] Margaret M Stewart and F Harvey Pough. Population density of tropical forest frogs: relation to retreat sites. *Science*, 221(4610):570–572, 1983. (Citado en la página 20).
- [40] Dan Stowell. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, 10:e13152, 2022. (Citado en la página 18).
- [41] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012. (Citado en la página 22).
- [42] Gasper Tkacik, Elad Schneidman, Michael J Berry II, and William Bialek. Ising models for networks of real neurons. *arXiv preprint q-bio/0611072*, 2006. (Citado en la página 22).
- [43] Daniel S Townsend and Margaret M Stewart. Reproductive ecology of the puerto rican frog *eleutherodactylus coqui*. *Journal of Herpetology*, pages 34–40, 1994. (Citado en la página 20).
- [44] Michael Towsey, Jason Wimmer, Ian Williamson, and Paul Roe. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecological Informatics*, 21:110–119, 2014. (Citado en la página 17).
- [45] Pauli Virtanen, Ralf Gommers, Travis Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stefan van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İbrahim Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Damien Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, José Quintero, Charles R. Harris, Anne M. Archibald, António H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. *Scipy sig-*

- nal processing (scipy.signal). <https://docs.scipy.org/doc/scipy/reference/signal.html>, 2020.
- [46] Lawrence L Woolbright. Patterns of nocturnal movement and calling by the tropical frog *eleutherodactylus coqui*. *Herpetologica*, pages 1–9, 1985. (Citado en la página 20).
- [47] Jie Xie, Towsey Michael, Jinglan Zhang, and Paul Roe. Detecting frog calling activity based on acoustic event detection and multi-label learning. *Procedia Computer Science*, 80:627–638, 2016. (Citado en la página 19).
- [48] Jie Xie, Mingying Zhu, Kai Hu, Jinglan Zhang, Harry Hines, and Ya Guo. Frog calling activity detection using lightweight cnn with multi-view spectrogram: A case study on kroombit tinkler frog. *Machine Learning with Applications*, 7:100202, 2022.
- [49] Hong-Li Zeng, Mikko Alava, Erik Aurell, John Hertz, and Yasser Rou-di. Maximum likelihood reconstruction for ising models with asynchronous updates. *Physical review letters*, 110(21):210601, 2013. (Citado en las páginas 22 y 35).
- [50] Tao Zhang, Guoqing Feng, Jinhua Liang, and Tong An. Acoustic scene classification based on mel spectrogram decomposition and model merging. *Applied Acoustics*, 182:108258, 2021. (Citado en las páginas 19 y 26).