

# Identificando Elenas (*Eleutherodactylus eileenae*) a partir de audios desordenados

Daniel Machado Pérez

Grupo C411

[DANIEL.MACHADO@GMAIL.COM](mailto:DANIEL.MACHADO@GMAIL.COM)

**Tutor(es):**

Dr. Roberto Mulet, *Facultad de Física*

## Resumen

Este estudio analiza las secuencias de cantos de la rana *Eleutherodactylus eileenae* (que llamaremos Elena), obtenidos mediante grabaciones de campo realizadas con nueve micrófonos distribuidos alrededor de los especímenes. Para identificar las secuencias probables de los cantos se propone un método que emplea mel-espectrogramas y análisis de las energías temporales de las grabaciones. Para garantizar la sincronización de los audios, se procedió con la obtención de los momentos de pico de energía y el cálculo de los desfases con correlación cruzada utilizando un archivo de referencia. Cada canto es conocido comúnmente como Colín, en imitación al sonido producido, en el que se emite una señal con frecuencia baja que denominaremos CO y una con frecuencia alta que llamaremos LIN. Los resultados revelan patrones en la estructura del coro, donde cada rana parece mantener una periodicidad consistente entre cantos y ajustar las frecuencias en respuesta a otros individuos.

## Abstract

This study analyzes the call sequences of the frog *Eleutherodactylus eileenae* (which we will refer to as Elena), obtained through field recordings made with nine microphones placed around the specimens. To identify the probable call sequences, a method is proposed that uses mel-spectrograms and temporal energy analysis of the recordings. To ensure the synchronization of the audios, the energy peak moments were obtained, and the time lags were calculated using cross-correlation with a reference file. Each call is commonly known as Colín, imitating the sound produced, where a signal with a low frequency, called CO, and a high frequency, called LIN, is emitted. The results reveal patterns in the structure of the chorus, where each frog seems to maintain a consistent period between calls and adjust its frequencies in response to other individuals.

**Palabras Clave:** Elena, Energía Temporal, Frecuencia, Mel-Espectrograma, Colín, Correlación Cruzada.

**Tema:** Sistemas Computacionales, Procesamiento y Análisis de Señales.

## 1. Resumen Extendido

### Breve Resumen del Estado del Arte

El estudio de las señales acústicas en animales para distinguir individuos y analizar su comportamiento ha evolucionado considerablemente. Blumstein et al. (2011) propusieron el uso de arreglos de micrófonos para la detección y localización de fuentes acústicas, destacando el potencial de las correlaciones cruzadas para diferenciar individuos. En el campo de la eco-acústica, Farina (2018) introdujo técnicas y métricas para caracterizar estructuras de coro y patrones de comportamiento, abriendo nuevas áreas de análisis en bioacústica. Akmentins (2011, 2015) implementó dispositivos de grabación automatizados y técnicas de clasificación mediante transformadas de Fourier para identificar especies y tipos de vocalización en ranas. En años recientes, se han desarrollado métodos de clasificación de coros mediante espectrogramas, como

los trabajos de Xie et al. (2022) y Gan et al. (2020), quienes emplearon redes neuronales y algoritmos de aprendizaje automático para detectar coros en condiciones ambientales diversas. Recientemente, investigaciones como las de Calsbeek et al. (2022) y Sánchez-Gendriz y Padovese (2017) han explorado las dinámicas de coro y la elección sexual, logrando avances en la caracterización individual de los sonidos en especies de ranas mediante análisis de niveles de presión sonora y técnicas de filtrado. Sin embargo, hasta la fecha, no se ha propuesto un método automático para discriminar individuos dentro de un coro utilizando exclusivamente arreglos de micrófonos.

### Problema Tratado

La rana *Eleutherodactylus eileenae* (Elena) es una especie endémica de Cuba. Su canto (Colín) es característico y se compone de dos partes: una señal con frecuencia baja que denominaremos CO y una

con frecuencia alta que llamaremos LIN. Estas señales se emiten en secuencias que se repiten a lo largo del tiempo, formando un coro.

La Elena es una especie estudiada por profesores e investigadores de la Facultad de Biología de la Universidad de La Habana. Como parte de sus investigaciones, fue planteada la necesidad de analizar el comportamiento de los coros antes mencionados. Se quiere saber si poseen una estructura determinada, si existen individuos que juegan un papel protagónico, entre otras cosas. Hasta el momento, el procesamiento de la información recopilada se realizaba manualmente, por lo que, para provocar avances más eficientes y rigurosos en los estudios, se impone la automatización de estos procesos y su trato con un enfoque más computacional.

El objetivo de este estudio es estudiar dichos coros, por lo que la primera tarea que se planteó fue la de identificar las secuencias probables de los cantos de Elena a partir de grabaciones de campo realizadas con nueve micrófonos distribuidos alrededor de los especímenes. Como los micrófonos están ubicados relativamente cercanos, cada uno capta los cantos de más de una rana. De ahí surge la necesidad de un método para distinguir los cantos del espécimen más cercano a cada dispositivo para, una vez hecho esto, conformar las secuencias generales. Para visualizar la distribución geográfica de los micrófonos véase la Figura 1

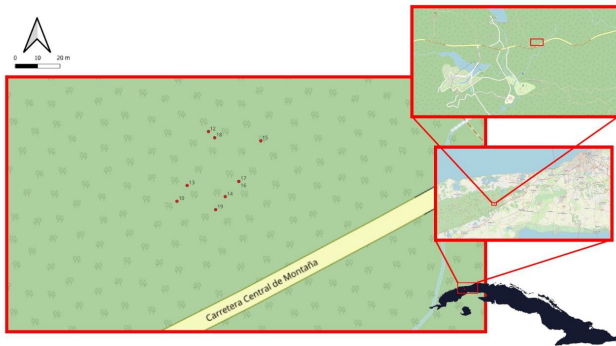


Figura 1: Distribución Geográfica de los Micrófonos.

Se cuenta con el dataset de las grabaciones hechas por los 9 micrófonos. Estas se realizaron durante 3 días, a partir de las 18:00 hasta las 06:00 horas del día siguiente, donde de cada hora se registraron 58 minutos y se descansaron 2. Los dispositivos se activaron remotamente y al mismo tiempo. Para lo que se expone en el presente informe, se utilizó la información captada el día 21 de octubre de 2023 a las 19:00 horas. En gran parte de la investigación se trabajará cada audio como el mel-espectrograma correspondiente. Los parámetros utilizados para la obtención de cada uno fueron los siguientes:

- hop\_length: 512
- n\_fft: 6096

- n\_mels: 128
- f\_min: 1600
- f\_max: 4096
- sample\_rate: 96000

Para mayor facilidad en el tratamiento de los archivos y su fácil distinción, fueron renombrados para asignarle a cada uno de los 9 audios una letra de la **a** a la **i**. En la Figura 2 se puede observar un ejemplo de mel-espectrograma de un fragmento del audio **h** a la hora y el día seleccionados.

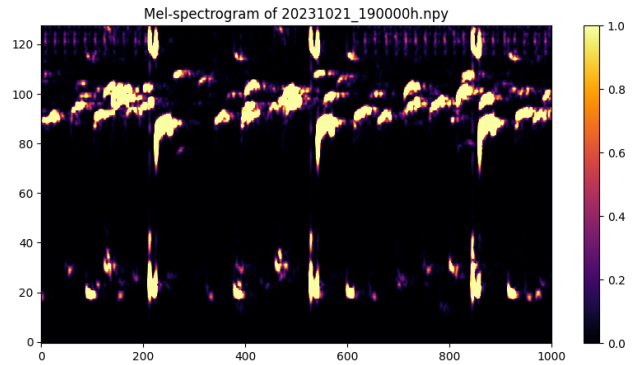


Figura 2: Ejemplo de Mel-Espectrograma de Fragmento del Audio **h**.

Previo al estudio que fue propuesto, es necesario preprocesar los archivos de audio. Fundamentalmente se debe garantizar un alto grado de sincronización. A pesar de que las grabaciones comenzaron a la misma hora, existe un *lag* no despreciable entre ellas, cosa que perjudica el análisis en cuestión. Para corregir dicho error se diseñó un proceso semi-automático que se sustenta en el método conocido de **Correlación Cruzada**. El plan consiste en la alineación de los archivos con base en un pivote que debe ser seleccionado. Esto se lleva a cabo mediante el cálculo de las energías temporales a partir de los mel-espectrogramas, la identificación de los momentos de pico de energía, la aplicación de **Correlación Cruzada** y la corrección de los desfases.

El proceso detallado de sincronización sigue los siguientes pasos:

#### Selección de un rango de bandas de frecuencia y un pivote.

Para ello se recomienda identificar un archivo en el que sea distinguible una señal (ya sea de CO o de LIN) que resalte en cuanto a intensidad, presencia en todos los archivos, y 'limpieza', queriendo indicar con esta último aspecto que en el rango de bandas de frecuencia en que se encuentra no exista un elevado número de otras señales que puedan constituir ruido y 'ensuciar' la información. El audio donde dicha señal se manifiesta con mayor claridad y/o intensidad puede ser seleccionado como pivote. Mientras tanto el

rango de bandas de frecuencia que contenga la señal y minimice el ruido puede ser el indicado para la sincronización. A partir de esa selección, el proceso restante se realiza con la información filtrada.

Por ejemplo, en el dataset que se usará para mostrar los resultados de este trabajo, se seleccionó como pivote al audio **h** y el rango de bandas de frecuencia de 113 a 117, por las características distintivas que se observan en la Figura 2.

### Cálculo de las energías temporales de cada archivo.

En cada instante de tiempo, la energía se calcula de la siguiente forma:

$$E_j = \sum_{i=0} a_{ij}^2, \forall j$$

donde:

$E_j$  es la energía en el instante de tiempo  $j$ ,

$a_{ij}$  es la intensidad en el instante de tiempo  $j$  y la banda de frecuencia  $i$ .

Para el audio **h**, su energía temporal se puede visualizar en la gráfica de la Figura 3

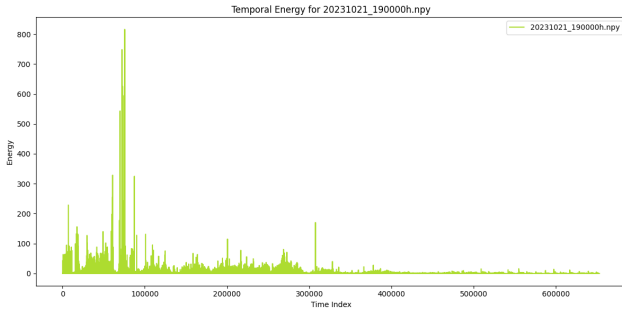


Figura 3: Energía Temporal del Audio **h**.

### Preprocesado de los archivos de audio aplicando cortes de frecuencia.

Para eliminar las bajas energías en cada audio, es decir, lo que se consideró ruido, se analizan los histogramas de las energías (Figura 4), y se calcula el ancho del primer bin. Dicho ancho será el mínimo de energía que se tomará en cuenta para filtrar la información relevante.

### Identificación de picos de energía en las señales.

La obtención de los momentos de pico de energía se lleva a cabo utilizando la función `find_peaks` de la biblioteca `scipy.signal`. Esta devuelve los índices en los que identifica un pico en la función que se le pasa. En este caso se le pasa la función de energías temporales de cada audio una vez eliminado el ruido.

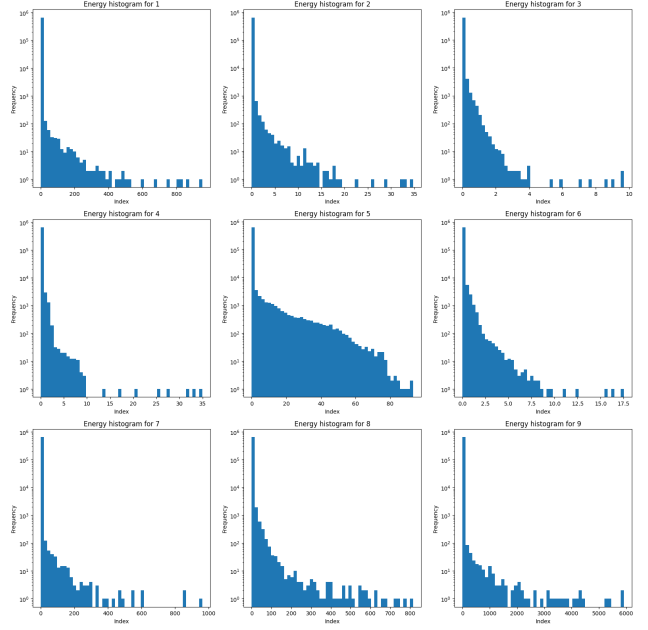


Figura 4: Histogramas de Frecuencias de las Energías Temporales de los Audios Filtrados por el Rango de Bandas Seleccionado.

El resultado es como el que se muestra en la Figura 5.

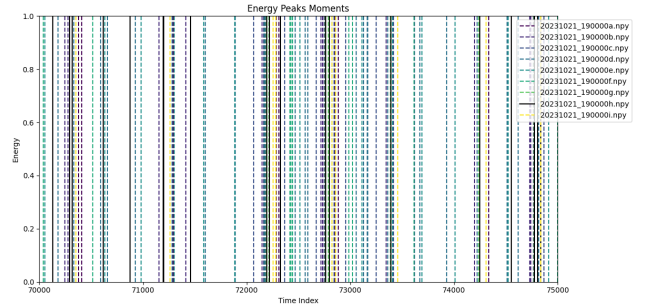


Figura 5: Momentos de Pico de Energía para un Fragmento de cada Audio.

### Cálculo de los desfases (*lags*) entre los picos de los archivos y el archivo pivote mediante Correlación Cruzada.

Luego se procede a calcular los desfases. Este método se basa en la suposición de que en el rango de frecuencias en el que una señal que sea captada por todos los audios de forma clara se debe experimentar un máximo de energía en el mismo momento, por lo que sincronizar los momentos de pico de energía es equivalente a sincronizar toda la información de los audios. La **Correlación Cruzada** aplicada a un par de *arrays* binarios calcula un coeficiente de coincidencias entre ambos, por lo que el *lag* se calculó como el desfase que maximiza dicha coincidencia. El resultado se ve como en la Figura 6.

### Aplicación de los desfases a los audios originales para su sincronización.

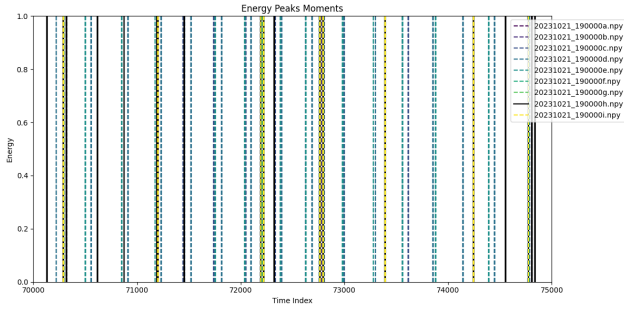


Figura 6: Momentos de Pico de Energía para un Fragmento de cada Audio después de Alinear.

Finalmente se aplican los desfases obtenidos a los archivos originales para comenzar con su procesamiento para la obtención de secuencias de cantos.

### Obtención de secuencias de cantos

Para llevar a cabo esta parte del estudio, se consideró oportuno separar la información de los archivos en CO y LIN, para realizar en paralelo los procesamiento en los momentos de baja y alta frecuencia respectivamente. De forma análoga a como se hace en la sincronización, se calculan las energías temporales y se establece un umbral de energía para eliminar el ruido en cada audio. El método propuesto para identificar el individuo más cercano a cada micrófono 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 las energías grandes y se elimina la información de esos cantos de las demás grabaciones, se obtendrán por separado los datos que se buscan. El algoritmo que representa lo antes expuesto consiste en lo siguiente:

- Se guardan copias vacías (llenas de ceros) de los *arrays* de energías temporales.
- Se selecciona de forma aleatoria uno de los 9 archivos.
  - En él se selecciona el índice donde se registra la energía máxima.
  - Si la energía es mayor que el umbral para el archivo seleccionado, se copia su valor en el índice del archivo vacío correspondiente.
  - Se borra el valor que exista en dicho índice en los 9 archivos (se coloca un 0).
- Se repite el proceso anterior hasta que en cada *array* de energía temporal solo queden valores por debajo de su respectivo umbral.
- Finalmente se guardan los resultados.

Este proceso se repite una cantidad determinada de veces para comparar los resultados de corridas diferentes y comprobar su consistencia.

Luego de hecho esto, el resultado de la identificación de Elenas por micrófonos, visto como una secuencia general, se vería como en la Figura 7

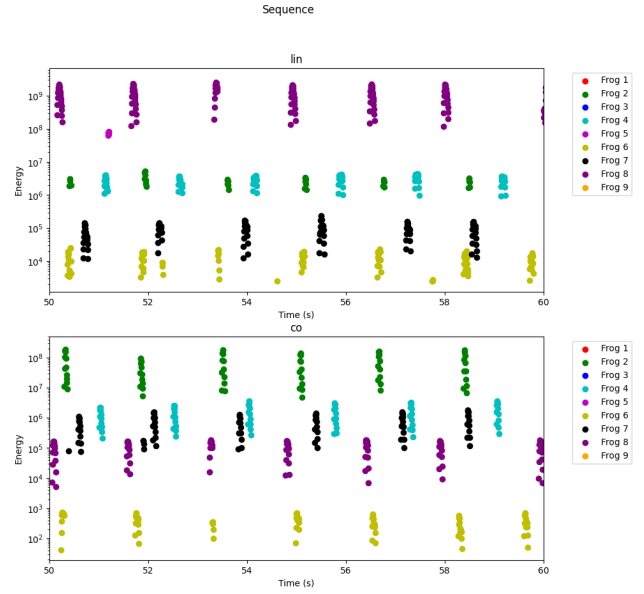


Figura 7: Fragmento de Secuencia de Cantos de las Elenas.

Una vez identificada la secuencia de cantos de cada Elena, se analizó el comportamiento del coro en cuanto a las bandas de frecuencia. De lo observado en las figuras 8 y 9 se puede formular la hipótesis de que las Elenas regulan la frecuencia de sus cantos en dependencia de las frecuencias de los demás miembros del coro, buscando baja superposición y tal vez mayor distinguibilidad.

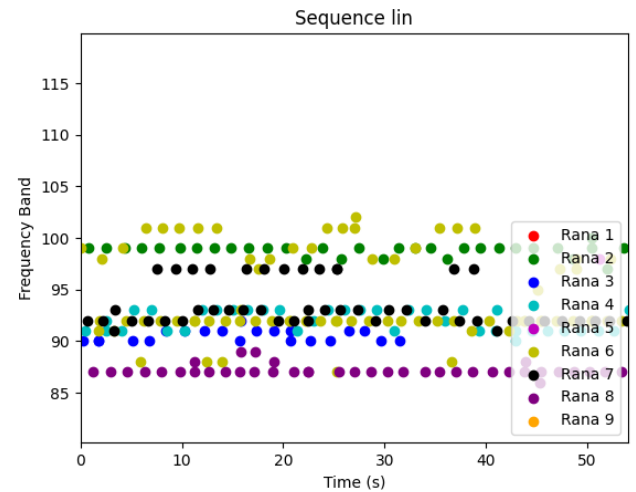


Figura 8: Comportamiento del Coro en Cuanto a la Frecuencia de los LIN.

El análisis de dicha hipótesis, la estadística de los coros, un estudio de causalidad, son algunas de las líneas futuras de esta investigación. El código fuente de este trabajo puede ser encontrado en el siguiente repositorio de GitHub: [Identifying Elenas](#). [JCE MatCom](#)

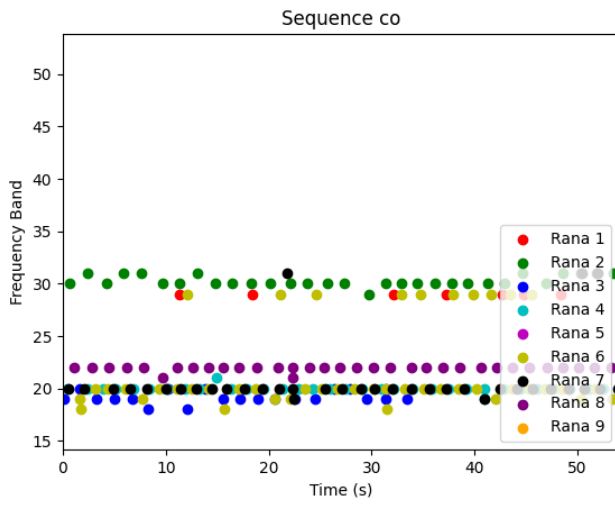


Figura 9: Comportamiento del Coro en Cuanto a la Frecuencia de los CO.