

Identificación de sonidos y la transformadas de Fourier

Angel Blanco*, M.Tarazona-Alvarado,[†] Yerimi Gamboa[‡]
Universidad Industrial de Santander, Bucaramanga, Colombia.

Índice

1. Introducción	1
2. Metodología	2
2.1. Python	2
3. El experimento y los resultados	3
4. Conclusiones y recomendaciones	7

Resumen

La transformada discreta de Fourier es una herramienta ampliamente utilizada para elaborar un análisis frecuencial de una señal, pero por sí sola no es suficiente para caracterizar cierto tipo de señales (cantos de aves, distinguir diferentes tipos de instrumentos o voces de personas diferentes). En el siguiente reporte se muestra una forma de caracterizar los cantos de *Paco*, un ave presuntamente perteneciente a la especie *Brotoferis jugularis*. Se usó como referencia los audios de una base de datos de esta especie y la especie *Brotoferis chiriri*. Gracias a nuestro método y por medio de una función de correlación se logró encontrar un patrón, el cual permite caracterizarlo. Se concluyó que existe una mayor similitud entre *Paco* y *Brotoferis jugularis*. Finalmente consideramos que está es una buena forma de analizar este tipo de datos, que se puede expandir a diferentes tipos de fenomenologías sonoras.

1. Introducción

El identificar a alguien entre cierto grupo de individuos ha sido de vital importancia a través de la historia, dicha identidad permite obtener rasgos que lo diferencia o asocia con otros individuos. La información puede ser común o única; si decimos que es común, entonces son las características que le hacen pertenecer a un algo, y si es única es porque la información o característica le identifica como un ser único. Un ejemplo de ello son las huellas dactilares en los seres humanos o la nariz de los perros; pero existe cierta característica que es más familiar, con la cual todos los seres interactúan e identifican a cada momento, el sonido.

El sonido que los seres vivos emiten es otra peculiaridad que les hace ser únicos y relacionarse con los de su propia especie. Debido a que cada sonido emitido, por ejemplo, la voz de una persona, tiene asociado unos niveles de potencia que son la huella dactilar del sonido que el emite; siendo este único. Un ejemplo muy palpable en el ser humano es ¿por qué un bebé recién nacido reconoce a su madre entre muchas otras personas? El bebé le identifica porque tiene asociado y reconocidos los niveles de potencia del sonido que tiene su mamá.

Nuestro principio está basado en esta unicidad de algunos sonidos, se propone identificar a un ave casera que posiblemente pertenece a la especie *brotoferis jugularis*, detectando sus niveles de potencia

*angelblanco43@gmail.com

[†]miguelta281@gmail.com

[‡]jeremi0112@gmail.com

asociados al sonidos que el emite. Se comparará con dicha especie y con *brotogetis chiriri* a partir de un análisis de señales utilizando la transformada de Fourier y herramientas de programación como lo son Python y MATLAB.

2. Metodología

Lo que hicimos primeramente consistió en recolectar los audios de nuestra ave, dichos audios se grabaron en formato .wav, ya que dicho formato ofrece un audio digital sin compresiones ni filtros que puedan alterar la calidad del mismo.

De manera similar, se buscó una colección de grabaciones online (*Xeno-canto*). Se buscaron audios de la especie *Brotogetis jugularis* y *Brotogetis chiriri*, que a su vez hubiesen sido grabados en la misma zona geográfica, ya que este es uno de los factores a considerar a la hora de caracterizar diferentes tipos de aves, debido a que se pueden presentar pequeñas diferencias entre mismas especies por pertenecer a lugares geográficos diferentes [2].

Para un correcto muestreo es necesario utilizar el criterio de Nyquist. Establece que se puede reconstruir una onda siempre que la frecuencia que se elija para muestrear sea mayor que el doble de la frecuencia mas alta encontrada en la señal, es decir

$$f_n = \frac{1}{2\Delta t}, \quad (1)$$

esto garantiza que la señal no está siendo sobre-muestreada. Lo anterior evita la identificación errónea de una frecuencia de señal, que puede introducir errores en el muestreo.

El siguiente procedimiento esta estructurado de tal forma que no solo permite analizar el canto de aves, también puede ser usado para el análisis de notas musicales, ladridos de perros e incluso voces. Es posible ampliarlo según la necesidad sin perder la estructura.

2.1. Python

Ya teniendo los diferentes audios en formato .wav, se hace un filtrado para determinar si están en *estéreo* o *mono*. De ser estéreo se separan los dos canales y se toma uno de ellos para continuar con el proceso. Esta elección es arbitraria pues no se encontraron mayores diferencias que favorezcan la elección.

Ahora aplicando la (Short-time Fourier transform) **STFT** la cual divide la señal en ventanas para procesar un numero limitado de muestras a la vez. La elección de esta ventana es importante, pues una ventana grande en el dominio temporal implica una ventana pequeña en el dominio frecuencial y viceversa. Para garantizar un análisis correcto de la señal se recurre al criterio de *Nyquist* (ecuación 1).

Para obtener el espectrograma en Python ya existen funciones definidas, entre ellas encontramos: *Matplotlib.pyplot.specgram()*, *scipy.signal.spectrogram()* y *librosa.core.stft()*. Debido a las características en la elección de la ventana y la forma de los valores de salida elegimos *matplotlib.pyplot.specgram()*.

Para aislar las frecuencias fundamentales usamos como criterio la potencia. Debido a que, aun siendo la misma especie de ave, no se tienen los mismos registros de potencia debido a las condiciones de grabación. Se hizo necesaria entonces una normalización que facilite la comparación de los audios. La normalización se llevo a cabo dividiendo por el máximo valor de potencia.

De esta forma tenemos las frecuencias fundamentales que componen esa señal (el canto de un ave) y su importancia dada por la potencia que alcanza en dicha frecuencia, con lo cual ya es posible comparar la similitud entre dos audios.

3. El experimento y los resultados

Los espectrogramas representan la energía del contenido frecuencial de una señal según varía en el tiempo.

La visualización de los espectrogramas (figura 1) otorga un criterio de potencia, ya que permite identificar las frecuencias fundamentales que conforman la señal (canto del ave).

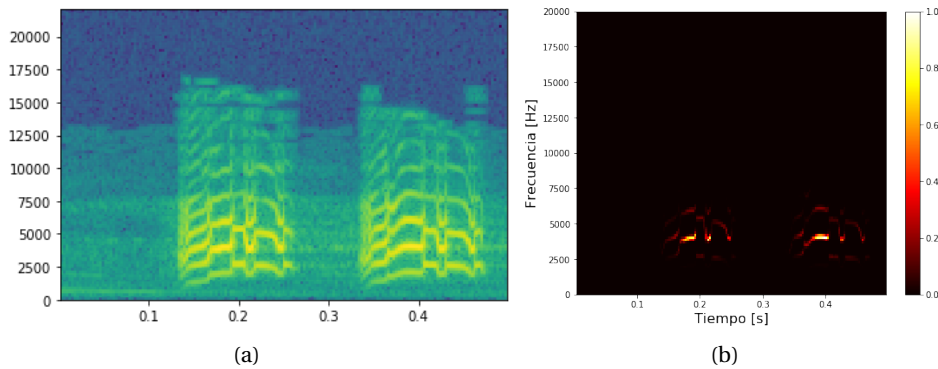


Figura 1: En la imagen (a) se muestra el espectrograma del audio obtenido de *Xeno-canto*. Corresponde a la especie *Brotogeris jugularis*. Esta figura nos permite identificar patrones característicos de la especie en el dominio frecuencial y temporal. Es evidente la presencia de ruido en la señal, para identificar con mayor facilidad las frecuencias fundamentales se procedió a normalizar (normalizada por el valor máximo). Esto se muestra en la imagen (b), en la cual el ruido pierde importancia evidenciando con mayor facilidad los patrones

A partir de esto y tomando como criterio la intensidad de la señal (potencia) se individualizan las frecuencias fundamentales. Gracias a los espectrogramas normalizados elegimos un valor de **0.2**. Este valor es superior al ruido presente en el audio, de esta forma es posible identificar las frecuencias más relevantes contenidas en el espectrograma.

Debido a que solo se está caracterizando el canto del ave, independientemente de si corresponde a una sílaba, canción o grito aleatorio [1]. Nos enfocamos únicamente en las frecuencias características (figura 2), dejando de lado el instante en el que son más relevantes.

Esto permite caracterizar los individuos de una especie y posteriormente relacionarlo con el de otras. Cabe resaltar que, si se desea identificar el tipo de canto, se deben tener en cuenta los instantes en los que las frecuencias alcanzan determinados valores de potencia.

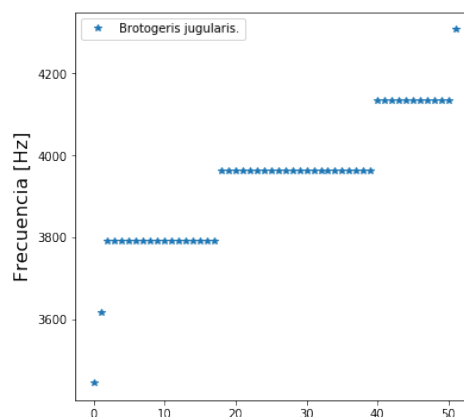


Figura 2: La figura muestra las frecuencias fundamentales de un *Brotogeris jugularis*. Corresponden a las frecuencias aisladas de la figura 9a. Se hace evidente que existen frecuencias muy puntuales que forman una huella para el ave. Recuerde que **STFT** se basa en el uso de ventanas temporales (y frecuenciales), por tanto, la existencia de varios registros que corresponden a la misma frecuencia indican que se identificó la misma frecuencia en diferentes ventanas temporales.

Como hemos mostrado, cada una de estas frecuencias aporta en diferentes niveles a la señal inicial. A las frecuencias se les asocia la mayor potencia que alcanzaron, esto como un indicativo de su importancia. Esto se puede visualizar en las figuras (3). El hecho de asociar a cada frecuencia fundamental una potencia define la *huella* del ave, con la cual ahora si podremos hacer comparaciones y caracterizar. De manera análoga esto sirve para diferencias diferentes tipos de instrumentos, ladridos de perros, voces de personas entre otras.

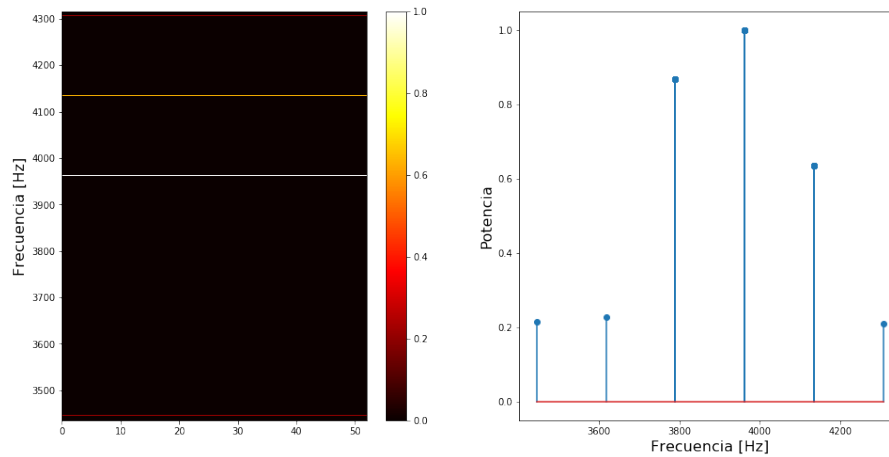


Figura 3: La figura muestra las frecuencias características y sus potencias asociadas de tal forma que se vea indicada su importancia en la señal. En la izquierda se muestra la importancia de la frecuencia en escala de colores. Se evidencia que no todas las frecuencias se alcanzan a apreciar (izq) debido a que los niveles de potencia son muy bajos. Una visualización mas completa se muestra en la figura de la derecha, donde si se pueden apreciar todas las frecuencias.

Utilizando el criterio mostrado en la figura (3) analizaran dos audios distintos de *Paco* tomados en días diferentes. Con el objetivo de tener un buen punto de partida que permita identificar y relacionar a *Paco* con una especie. Y evidenciar la validez de nuestro método a la hora de caracterizar sonidos.

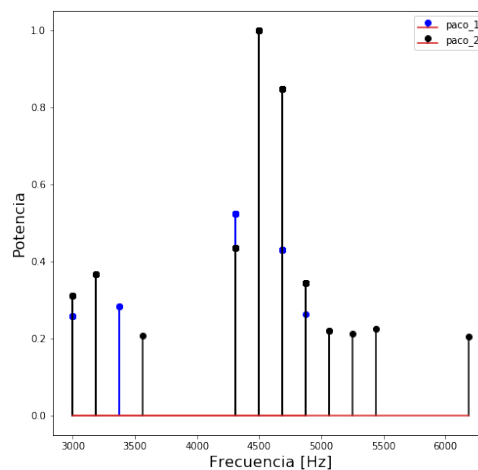


Figura 4: En la figura se muestra la superposición de la *Huella* dos audios pertenecientes a *Paco*. Se evidencia que comparten la frecuencia principal. Comparten además casi todas las frecuencias fundamentales.

En la figura (4) se muestra la semejanza que existe entre los dos audios. Se tomará esto como el punto de partida para garantizar la eficacia del método. dos huellas de diferentes sonidos se mostrará la similitud que existe entre las aves.

Entre mayor sean las frecuencias en común a la hora de superponer las dos huellas de diferente sonido, mayor será la relación entre los mismos.

Ahora, se aplicarán los criterios descritos anteriormente para caracterizar a nuestra ave *Paco* la cual

pertenece posiblemente a la especie *Brotogeris jugularis*. Se comparará con otro espécimen de su misma especie y con otra ave perteneciente a la especie *Brotogeris chiriri*.

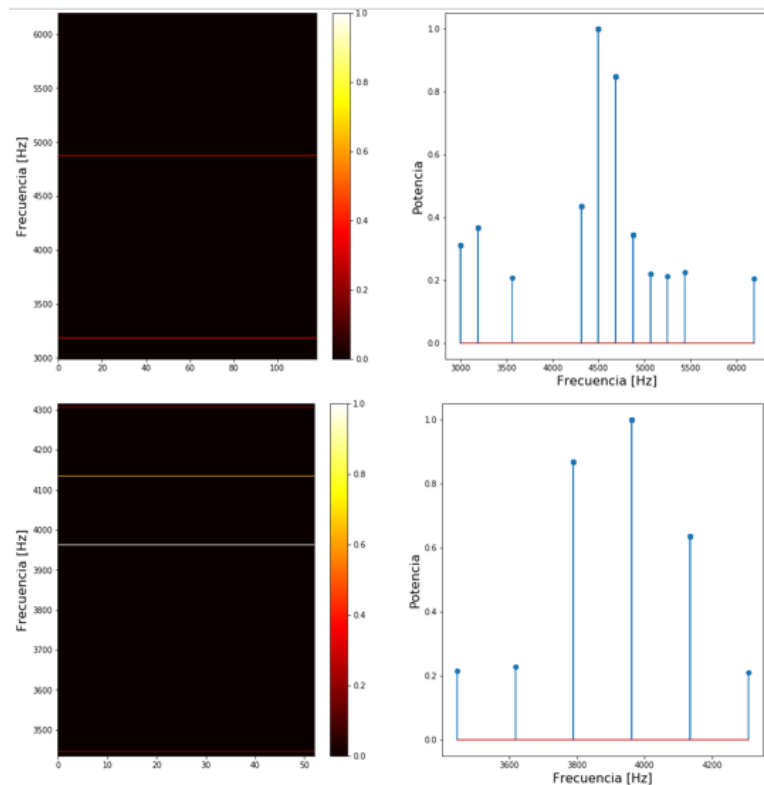


Figura 5: De manera similar a la figura (3) la figura muestra las frecuencias características de *Paco* (arriba) y *Brotogeris jugularis* (abajo). Se observa que las frecuencia que posee mayor potencia para *Paco* corresponde a 4500 Hz mientras que para *Brotogeris jugularis* corresponde a 3962 Hz.

La figura (5) muestra que no hay una coincidencia en las frecuencias fundamentales que caracterizan a *Paco* y a *Brotogeris jugularis*. Esto se hace mas evidente en la figura 6.

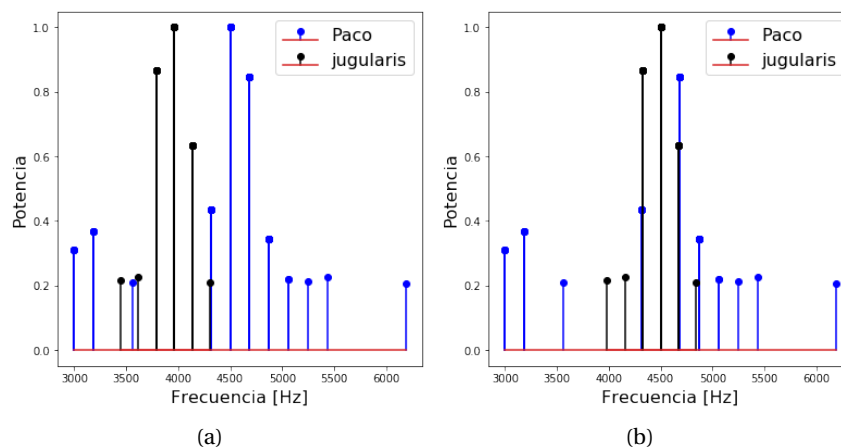


Figura 6: La figura (a) muestra la superposición de la *huella* de *Paco* y *Brotogeris jugularis*. Se observa que solo comparten una frecuencia. No se evidencia un patrón común en las dos huellas. La figura (b) corresponde a un corrimiento de *Brotogeris jugularis* en 537.89 Hz. Se ve que este arreglo muestra un patrón en cuatro de sus frecuencias fundamentales.

En la figura (6.a) queda explicito que *Paco* no pertenece a *Brotogeris jugularis*. El corrimiento de la figura (6.b) se hizo haciendo coincidir los fundamentales principales como punto de referencia para develar

algún patrón que compartan las dos aves. Claramente en la figura (6.b) comparten cuatro frecuencias fundamentales habiendo desplazado *Brotogeris jugularis* 537.89 Hz. No es de extrañar que gracias a este desplazamiento se comparta un segmento de la huella, ya que *Paco* no pertenece a la misma especie (Jugularis) pero si al mismo genero (*Brotogeris*)

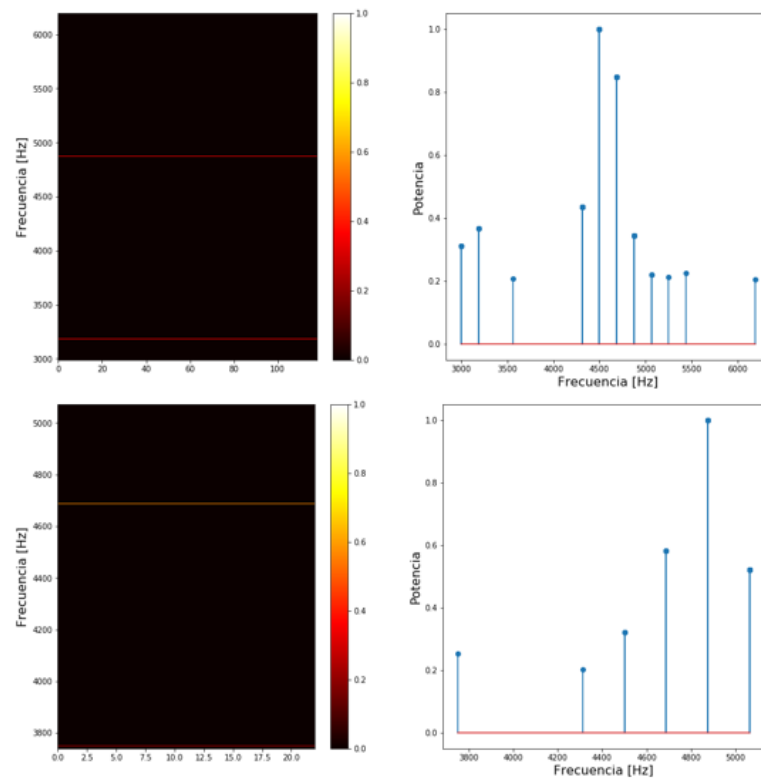


Figura 7: De manera similar a la figura (3) la figura muestra las frecuencias características de *Paco* (arriba) y *Brotogeris chiriri* (abajo). Se observa que las frecuencia que posee mayor potencia para *Paco* corresponde a 4500 Hz mientras que para *Brotogeris chiriri* corresponde a 4875 Hz.

Como se observa en la figura (7) muestra que existe un patrón de coincidencia en las frecuencias fundamentales que caracterizan a *Paco* y *Brotogeris chiriri*. Se puede posiblemente afirmar que *Paco* pertenece a la especie *chiriri*. Esto se evidencia mejor en la figura (8)

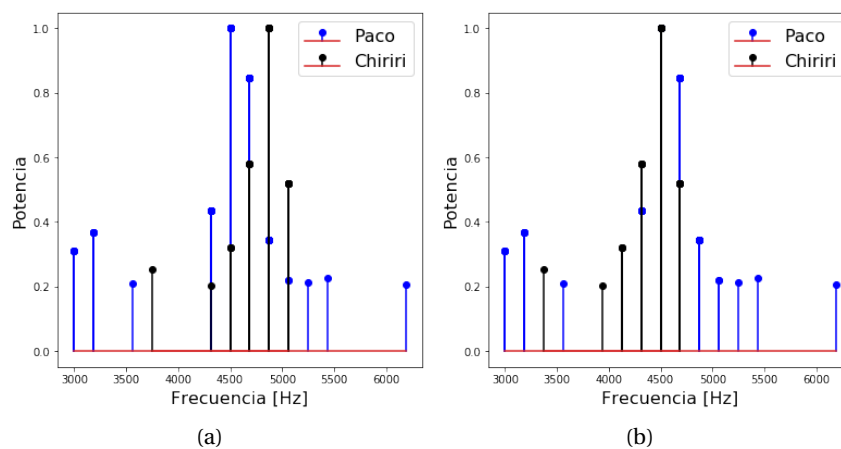


Figura 8: La figura (a) muestra la superposición de la *huella* de *Paco* y *Brotogeris chiriri*. Se evidencia que comparten cinco de sus frecuencias fundamentales. Se muestra que existe un patrón común que caracteriza a ambas señales. La figura (b) corresponde al corrimiento de *Brotogeris chiriri* en 375 Hz. Podemos observar que se de esta forma se superponen sus tres frecuencias principales.

Como se mostró en la figura (8.a) es bastante mas probable que *Paco* pertenezca a la especie *chiriri*, ya que sin necesidad de corrimientos coincide en cinco de sus frecuencias fundamentales. Se observa que existe un patrón común en las dos huellas. En la figura (8.b) se evidencia que gracias al corrimiento, las tres frecuencias principales coinciden. Lo que mas destaca de la figura (8.b) es que la frecuencia principal (4500 Hz) que tiene asociada la mayor potencia, coincide perfectamente.

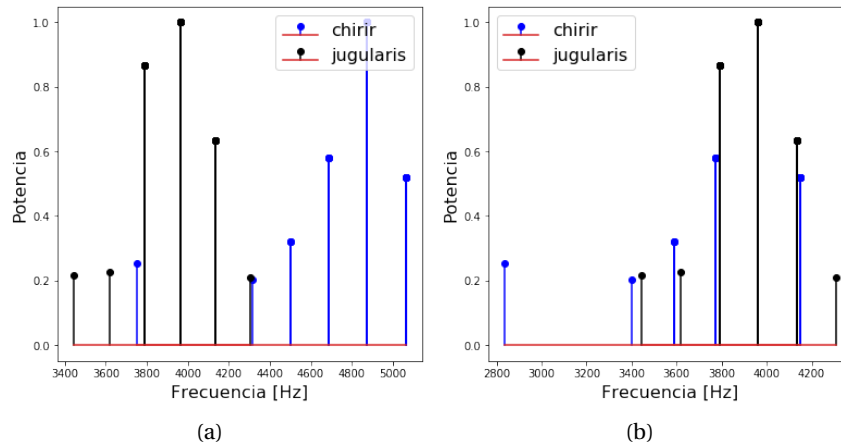


Figura 9: En la figura (a) se observa la superposición de las *huellas* de *Brotogetis jugularis* y *Brotogetis chiriri*. A priori se observa que no hay mucha similitud entre ellas. Solo coincide una de sus frecuencias y no es una de sus principales. La figura (b) corresponde al corrimiento de *Brotogetis jugularis* en 913 Hz. Se evidencia que existe cierta similitud en los patrones formados por las frecuencias fundamentales de cada *huella*.

En la figura (9.b) se muestra cierta similitud en las *huellas* de ambas especies. No es de extrañar que exista este grado de similitud ya que las dos especies pertenecen al mismo genero.

4. Conclusiones y recomendaciones

Es sorprendente como la transformada de Fourier permite analizar diferentes fenómenos. En el caso de señales de audios permite una visualizar el sonido a través de espectrogramas y espectros de potencia.

Como consecuencia de lo mostrado en el reporte se puede concluir que la *huella* de *Paco* es más similar a la especie *Brotogetis chiriri*.

Esto se visualiza en la figura (8.a), pues *Paco* comparte 5 de las frecuencias con *Chiriri*, aunque no con las mismas intensidades. Al realizar la superposición (figura 8.b) de la fundamental principal solo comparan tres frecuencias, las cuales corresponden a las tres fundamentales principales tanto de *Paco* como de *Chiriri*. Esto es contrario a lo que se esperaba inicialmente, que la mayor similitud fuese con la especie *Brotogetis jugularis* debido a comparaciones morfológicas.

Como era de esperarse, existe un grado de similitud entre *Jugularis* y *Chiriri*. Esto radica en la proximidad de ambas especies y se observa en la figura (??). Solo se ve luego de superponer las frecuencias principales. Se observa el mismo patrón al hacer un desplazamiento de 912 Hz

Gracias a los esfuerzos mostrados se logró caracterizar un ave hogareña. Se encontraron sus frecuencias fundamentales, así como su *huella* propia ligada a la potencia. El próximo esfuerzo radica en considerar el comportamiento de las frecuencias en las ventanas temporales. Esto permitirá mostrar cuando un sonido emitido por el ave corresponde a una canción, sílaba o grito aleatorio ([1]).

Se recomienda que, para caracterizar los cantos se analicen en que instantes y que duración tienen las frecuencias características. Por plantear un caso, podría ser relevante que luego de un intervalo de tiempo en el que se emiten frecuencias poco pronunciadas (poca potencia), haya un silencio y luego se emita la frecuencia más potente.

Una de las grandes fortalezas en como se abordó el fenómeno radica en que no solo se limita al mismo, es decir, es fácilmente aplicable a otros fenómenos (reconocimiento de voz, determinar diferencias entre instrumentos entre otros).

Sin duda una de las cosas mas satisfactorias de este proyecto es que permitió romper con creencias o hechos que se tomaban como ciertos. Este es el camino de las ciencias, que a base de evidencia y rigurosidad rompe con convicciones inherentes a nuestra ignorancia.

Referencias

- [1] Joseph A Kogan and Daniel Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. *The Journal of the Acoustical Society of America*, 103(4):2185–2196, 1998.
- [2] Carlos Alberto Maciel-Mata, Norma Manríquez-Morán, Pablo Octavio-Aguilar, and Gerardo Sánchez-Rojas. El área de distribución de las especies: revisión del concepto. *Acta universitaria*, 25(2):03–19, 2015.

[2] [1]