Monitoreo Bioacústico de Aves utilizando Aprendizaje no Supervisado

Bioacoustical Bird Monitoring using Unsupervised Learning

Kenji Contreras[*1], Andrew Hall[2], Jaime Huertas[3], Fernando Merchán[4], Alejandro Von Chong[5]*
[1]Licenciatura en Ingeniería en Telecomunicaciones, Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá
[2]Licenciatura en Ingeniería en Control y Automatización, Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá
[3]Licenciatura en Ingeniería en Control y Automatización, Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá
[4]Departamento de Ingeniería de Sistemas de Comunicación, Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá
[5]Departamento de Control e Instrumentación, Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá

Resumen:

Palabras clave: Vocalizaciones de aves, espectrogramas, funcion de autocorrelacion, PCA, agrupacion

Abstract:

Keywords: Bird vocalizations, spectrograms, autocorrelation function, PCA, clustering.

* Corresponding author: kenji.contreras@utp.ac.pa

1. Introducción

La bioacústica es una ciencia multidisciplinaria que estudia la producción, transmisión y recepción de sonidos de animales tales como aves y mamíferos. Una de sus aplicaciones más populares, el monitoreo bioacústico, es una metodología empleada por biólogos y ecologistas para evaluar el impacto de la interferencia humana en la biodiversidad y apoyar esfuerzos para la preservación de especies salvajes haciendo uso de los continuos avances en Aprendizaje Automatizado (Machine Learning o ML por sus siglas en ingles) para propósitos de procesamiento y análisis de datos [1].

Las aves juegan un papel muy importante en diferentes ecosistemas, contribuyendo al control de poblaciones de insectos y conservación de la flora dispersando semillas y polinizando plantas [4]. De tal manera, son un buen indicador para evaluar cambios en un hábitat debido a su distribución poblacional sobre un amplio rango de áreas y facilidad de detección en comparación con otros animales [5]. En este contexto, se ha visto la necesidad de poder procesar archivos de audio de una manera más eficiente debido al auge de nuevas tecnologías de grabación utilizadas por ecologistas para recolectar datos por periodos extendidos de tiempo, con fines de entrenar y optimizar modelos de ML cada vez más complejos orientados a la detección automática de aves [2].

1.1 Propiedades y análisis de vocalizaciones de aves

Una propiedad importante para considerar de cualquier sonido es si es estacionario o no, es decir, determinar si sus propiedades no cambian substancialmente a través del tiempo. Las vocalizaciones de las aves no son estacionarias debido a que tienen una corta duración y varían rápidamente [31]. Esta consideración impacta significativamente el tipo de estrategia

que se debe utilizar para detectar y extraer los sonidos deseados.

A diferencia de otras especies de animales, las vocalizaciones de aves están divididas en cuatro niveles jerárquicos: notas, silabas, frases y cantos [6]. En varias especies a nivel regional se presentan diferencias en términos de patrones de frases y cantos, indicando que las silabas, que se pueden considerar como un bloque elemental de una vocalización, son la fuente de información más adecuada para identificar automáticamente diferentes especies de aves.

Por ultimo, cabe destacar que analizar una señal no es un enfoque apto para tratar señales no estacionarias, por lo tanto, realizar cálculos en segmentos de pequeñas ventanas de tiempo es una practica común asumiendo que una señal es estacionaria durante una corta duración de tiempo.

Esta técnica, utilizando ventanas cuya función decaiga a cero el final de su respectivo rango con un determinado porcentaje de traslape entre ventanas, es la base de algoritmos como la Transformada de Fourier en Tiempo Corto (STFT por sus siglas en ingles) cuyo propósito es generar el espectrograma, una representación bidimensional de una señal en términos de tiempo y frecuencia

1.2 Métodos comunes de detección de vocalizaciones de aves

Algunos de los métodos más comunes se basan en la segmentación de silabas por medio de modelos estadísticos como el Modelo Oculto de Markov (HMM por sus siglas en inglés) [19], la detección por umbrales de energía [20] y el análisis de segmentos tiempo-frecuencia como la correlación cruzada de espectrogramas [22]. Es importante remarcar que estos métodos requieren la implementación de un esquema de supresión o filtrado de ruido para funcionar apropiadamente [21].

1.3 Extracción de características

Este procedimiento(Feature Extraction o por sus siglas en ingles) se puede definir como la reducción o simplificación de las características mas importantes que representan la información original para facilitar la etapa de análisis. Para experimentos de bioacústica, una señal de audio cruda no es apropiada para ser analizada debido a su alta dimensionalidad, causando que sonidos perceptualmente similares no puedan ubicarse en un espacio vectorial cercano [9].

El Análisis por Componente Principal (PCA por sus siglas en ingles) es un método que permite reducir la cantidad de información requerida para representar la información de entrada, ha sido empleado en trabajos de bioacústica para simplificar el análisis de espectrogramas [15, 25].

Vale la pena destacar que según [3], el pre-procesamiento del espectrograma puede mejorar la etapa de análisis. En este aspecto, se han empleado espectrogramas a diferentes escalas (lineales o logarítmicas), incluyendo la aplicación de binarización [16], este ultimo proceso permite eliminar segmentos residuales de ruido y preservar la información espectral mas significativos.

1.4 Modelos de Agrupación

Los modelos de agrupación o *clustering* se caracterizan por categorizar información de acuerdo con sus similitudes de manera no supervisada, es decir, calculan e infieren resultados automáticamente sin necesidad de una referencia previa.

Es pertinente mencionar que, en el ámbito de la bioacústica estos modelos se han usado como parte de metodologías de detección y censado de diferentes especies de animales tales como ballenas [24] y manatíes [25] a partir de la agrupación de espectrogramas.

En la literatura consultada referente a la detección automática de aves, el *clustering* se emplea como un método de Aprendizaje de Características (Feature Learning o FL por sus siglas en inglés), esto es, la transformación de información original a partir de sus características inherentes para mejorar el rendimiento de análisis posteriores [23, 26].

En este trabajo se realizo un estudio de rendimiento con diferentes modelos de *clustering* para determinar si un esquema de detección de aves basado en la agrupación de espectrogramas es lo suficientemente preciso para estimar la cantidad de especies de aves presentes en un determinado intervalo de tiempo y adicionalmente, si es posible detectar diferentes individuos de una misma especie con el mismo esquema.

En la sección de metodología se describe el esquema de detección de vocalizaciones, las técnicas de FE implementadas y el funcionamiento de los modelos de *clustering*. Por ultimo, en la sección de resultados se discuten los 2 experimentos realizados con el objetivo de sustentar la utilidad de la metodología de este proyecto para el censado bioacústico de aves.

2. Metodología

La programación e implementación de los algoritmos y modelos descritos en esta metodología se realizo en lenguaje Python debido a su alto nivel de optimización, conveniencia y disponibilidad de software de libre distribución orientado a ML y procesamiento digital de señales.

2.1 Recolección de datos

Para este trabajo se recolectaron grabaciones de audio del sitio web Xeno-canto [27]. Se seleccionaron 8 especies de aves comunes en Panamá (*Harpia harpyja, Crypturellus soui, Setophaga petechia, Hylophylax naevioides, Progne chalybea, Todirostrum cinereum, Xiphorhynchus susurran y Ramphastos sulfuratus)*, procurando escoger grabaciones con una calidad aceptable de audio a una frecuencia de muestreo de 44.1kHz.

2.2 Esquema de detección de vocalizaciones

2.2.1 Supresión de ruido

Esta primera etapa corresponde a la aplicación de un algoritmo capaz de minimizar la cantidad de ruido presente en las diferentes grabaciones. Para esta tarea se utilizó el Filtro Iterativo de Wiener, uno de los algoritmos mas usados para supresión de ruido en trabajos de bioacústica [21].

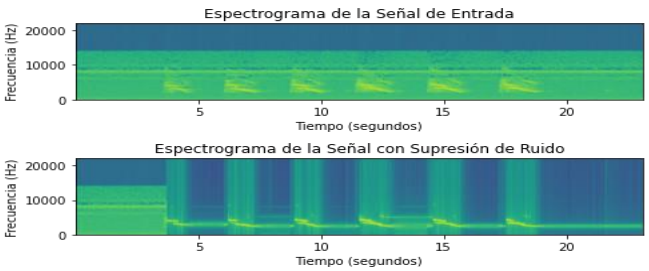
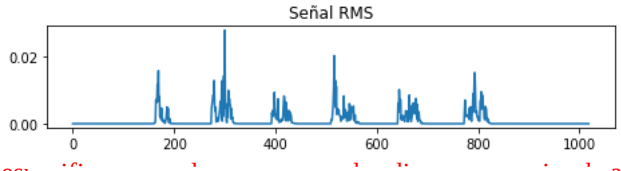


Figura 1. Supresión de ruido del canto de una Harpia.

A diferencia de otros algoritmos descritos en [21], este no requiere designar una señal de ruido para diferenciar las señales deseadas, el mismo algoritmo realiza una estimación del ruido por medio de un modelo autorregresivo conocido como LPC (Linear Predictive Coding por sus siglas en ingles).

2.2.2 Segmentación de silabas por medio de la Función de Autocorrelación

La función de autocorrelación (ACF por sus siglas en ingles) se ha utilizado para segmentar vocalizaciones de aves [20]. Sin embargo, en este trabajo se probo el método basado en [25, 30] debido a su efectividad para detectar componentes armónicos en diferentes sub-bandas de frecuencia y de tal manera segmentar una vocalización con características espectrales



es aves

Figura 2. Señal RMS producto del análisis por ventana de ACF.

El método consiste realizar un análisis por ventanas de 2000 muestras, y en cada una computar la ACF y su valor RMS utilizando el rango de 20 a 200 muestras. Este proceso iterativo genera una señal con picos de amplitud en los tramos donde se ubican las silabas (Figura 2).

La señal resultante es pasada por un filtro promedio para eliminar ruido transitorio (Figura 3) y posteriormente, se aplica un criterio de umbral basado en la mediana de la señal donde todo valor que supere el valor de n veces la mediana (donde el valor de n es arbitrario) es redondeado a 1, de lo contrario es 0 (Figura 4).

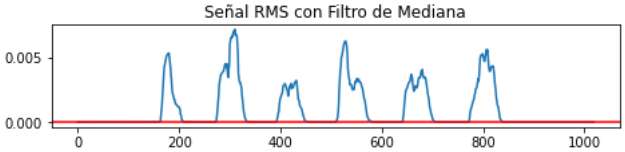


Figura 3. Señal resultante del filtrado por mediana.

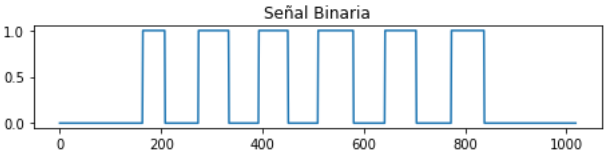


Figura 4. Redondeo de amplitudes de la señal RMS en base al criterio

Por ultimo, se realiza un análisis para detectar señales no deseadas que hayan superado el valor de umbral por medio de un criterio de duración (tomando en cuenta la consecución de muestras con valor unitario) para omitir señales de ruido de corta o muy larga duración.

Finalmente, se extraen las sílabas de la señal de entrada utilizando los intervalos descritos por los tramos de valores unitarios de la señal binaria.

2.3 Generación y Procesamiento de espectrogramas

Una vez obtenidas las silabas, se generaron los espectrogramas (Figura 5) utilizando una ventana de 1024 muestras (esto corresponde a la resolución en frecuencia de la STFT) con 50% de traslape, estos parámetros presentan un balance adecuado de resolución tiempo y frecuencia [9, 25].

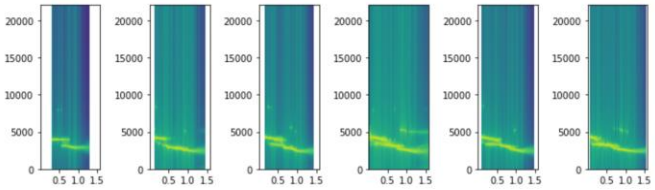


Figura 5. Espectrogramas en escala lineal.

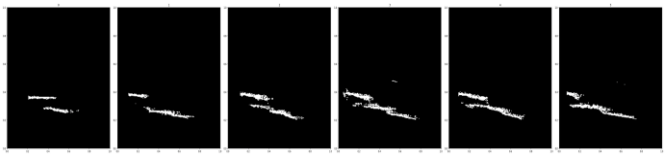


Figura 6. Espectrogramas binarios.

Posteriormente se realizo el proceso de binarización, este consiste en reducir a cero el valor normalizado de los pixeles que no estén por encima de n veces el valor promedio (donde n es un valor arbitrario) y redondear el resto a 1.

Por ultimo, se rellenaron las imágenes con ceros para obtener un set de espectrogramas con dimensiones uniformes de 257x150 pixeles (Figura 6).

2.4 Reducción de dimensionalidad por PCA

Una vez obtenidos los espectrogramas a partir del esquema de detección, se procede a la etapa de FE. Los n espectrogramas fueron convertidos en una matriz de n por m dimensiones donde m corresponde a la longitud resultante del aplanamiento de la imagen, es decir, la transformación de una matriz de dos dimensiones en un vector con longitud equivalente a la multiplicación de sus respectivas dimensiones.

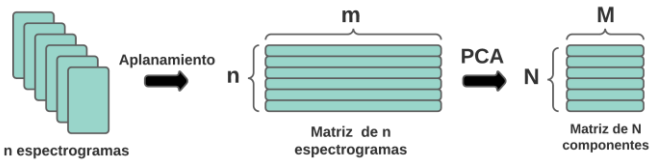


Figura 6. Visualización del proceso de reducción de dimensionalidad

El método de reducción de dimensionalidad por PCA consiste en proyectar las n variables a un sub-espacio ortogonal de manera que puedan ser representadas como un conjunto de N componentes no correlacionados, donde M equivale a los vectores propios que representan la información comprimida.
PCA tratara de comprimir la mayor cantidad de información en los primeros componentes, por lo tanto, solo se utilizarán los que contengan el mayor porcentaje de información.

2.5 Agrupación de espectrogramas

Si se pudieran visualizar en un espacio multidimensional, los vectores de la matriz de componentes que correspondan a espectrogramas similares estarían ubicados en un espacio vectorial cercano, por lo tanto, es posible estimar esta similitud en base al calculo de distancias con respecto a un punto de referencia.
En este contexto, es posible ejecutar la misma tarea de diversas formas, el factor determinante es el método de ejecución y por lo tanto, se diseñaron 2 experimentos con el objetivo de probar 5 modelos de clustering con diferentes funcionalidades.
El primer experimento consistió en determinar el mejor método de clustering que agrupar espectrogramas de diversas especies de aves. En el segundo experimento se aplico la misma metodología a la agrupación de espectrogramas de diferentes individuos de la misma especie, en este caso se escogió la *Harpia harpyja* como referencia debido a su interés como ave nacional de Panamá y su estatus de especie en peligro de extensión [34]

2.6 H*ard clustering vs soft clustering*

Los modelos de *clustering* realizan la estimación de agrupaciones de dos maneras, estas se pueden definir como *hard clustering* y *soft clustering* [33].
Agrupación por K-Medias (K-Means Clustering o KMC por sus siglas en ingles) y Agrupación Jerárquica (Hierarchical Clustering o HC) (Figura 7) categorizan muestras de manera que cada una solo puede pertenecer a una agrupación, estos son categorizados como métodos de *hard clustering.*

KMC buscara crear K agrupaciones en base al calculo iterativo de puntos de referencia (denominados centroides) con respecto a la distancia euclidiana de cada muestra. HC realizara una tarea similar pero basándose en el calculo de similitud entre pares de muestras, su funcionamiento se puede visualizar como un dendrograma, el cual ilustra las relaciones jerárquicas entre las muestras (Figura 7).
Adicionalmente, se incluyo una variante de KMC conocida como el algoritmo de Elkan [35], este algoritmo busca acelerar y optimizar el calculo de distancias aplicando el teorema de desigualdad triangular.

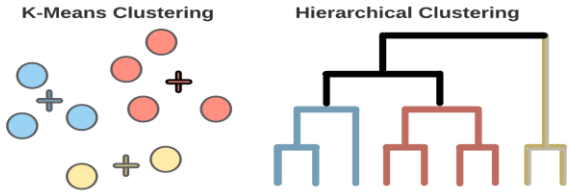


Figura 7. Ilustración simplificada de *hard clustering.*

Por otra parte, otros modelos como C-Medias Difusas (Fuzzy C-Means Clustering o FCM) y Modelos de Mezclas Gaussianas (Gaussian Mixture Models o GMM) (generan agrupaciones superpuestas donde cada muestra puede pertenecer a una o mas agrupaciones, estos métodos se categorizan como *soft clustering* (Figura 8).

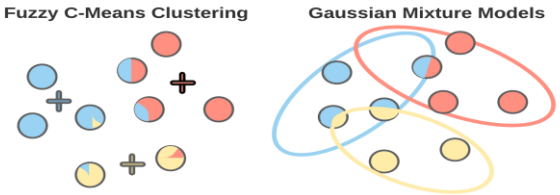
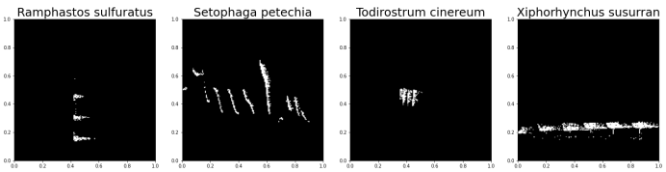


Figura 8. Ilustración simplificada de *soft clustering.*

FCM funciona de una manera inductiva como KMC, este en cambio generara agrupaciones superpuestas definidas por lógica difusa basada en el calculo de valores de membresía que determinaran le distribución de agrupación de cada muestra.
Por ultimo , GMM se considera como un método probabilístico que asume a existencia de una cantidad predefinida de distribuciones gaussiana (también llamadas mezclas) , para las cuales es necesario determinar sus parámetros y probabilidades desconocidas ,las cuales determinaran las respectivas agrupaciones de muestras.

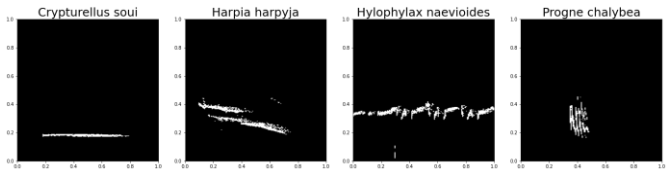
3. Resultados



Quisbert, Darlene Daiana

hace 8 meses

MUCHAS GACIAS PERO DEBERAS MIL GRACIAS! Con tu video entendí de una el tema con el cual tengo que hacer mis trabajos, seguí así capo!



Varianza 90%			
	H	C	VM
KMC	0.4318	0.6358	0.5143
KMC (Elkan)	0.6457	0.7719	0.7032
HC	0.5082	0.703	0.5899
FCM	0.2119	0.5622	0.3078
GM	0.4686	0.6737	0.5527

Varianza 80%			
	H	C	VM
KMC	0.4905	0.6799	0.5698
KMC (Elkan)	0.5254	0.716	0.606
HC	0.5991	0.7567	0.6688
FCM	0.4349	0.5705	0.4936
GM	0.4903	0.7056	0.5785

4. Conclusiones

AGRADECIMIENTOS

A la Secretaría Nacional de Ciencia y Tecnología y a la Universidad Tecnológica de Panamá, por abrir el espacio para el desarrollo y la presentación de proyectos innovadores e investigaciones provechosas para la comunidad.

REFERENCIAS

[1] J. Snaddon, G. Petrokofsky, P. Jepson, and K. J. Willis, "Biodiversitytechnologies: tools as change agents," 2013.

[2] D. Teixeira, M. Maron, and B. J. van Rensburg, "Bioacoustic monitoringof animal vocal behavior for conservation,"Conservation Science andPractice, vol. 1, no. 8, p. e72, 2019.

[3] P. Jancovic and M. K¨ok¨uer, "Bird species recognition using unsupervisedmodeling of individual vocalization elements,"IEEE/ACM Transactionson Audio, Speech, and Language Processing, vol. 27, no. 5, pp. 932–947, 2019.

[4] C.-H. Lee, S.-B. Hsu, J.-L. Shih, and C.-H. Chou, "Continuous birdsongrecognition using gaussian mixture modeling of image shape features,"IEEE Transactions on Multimedia, vol. 15, no. 2, pp. 454–464, 2012.

[5] C. K. Catchpole and P. J. Slater,Bird song: biological themes andvariations. Cambridge university press, 2003.

[6] D. Stowell and M. D. Plumbley, "Automatic large-scale classificationof bird sounds is strongly improved by unsupervised feature learning,"PeerJ, vol. 2, p. e488, 2014.

[7] P. Jancovic and M. K¨ok¨uer, "Bird species recognition using unsupervisedmodeling of individual vocalization elements,"IEEE/ACM Transactionson Audio, Speech, and Language Processing, vol. 27, no. 5, pp. 932–947, 2019.

[8] V. Roger, M. Bartcus, F. Chamroukhi, and H. Glotin, "Unsuper-vised bioacoustic segmentation by hierarchical dirichlet process hiddenmarkov model," inMultimedia tools and applications for environmental& biodiversity informatics. Springer, 2018, pp. 113–130.

[9] A. L. McIlraith and H. Card, "Birdsong recognition with dsp andneural networks," inIEEE WESCANEX 95. Communications, Power, andComputing. Conference Proceedings, vol. 2. IEEE, 1995, pp. 409–414.

[10] D. Stowell and M. D. Plumbley, "Audio-only bird classification usingunsupervised feature learning," 2014.

[11] X. Mankun, P. Xijian, L. Tianyun, and X. Mantian, "A new time-frequency spectrogram analysis of fh signals by image enhancementand mathematical morphology," inFourth International Conference onImage and Graphics (ICIG 2007). IEEE, 2007, pp. 610–615.

[12] J. Katz, S. D. Hafner, and T. Donovan, "Assessment of error rates in acoustic monitoring with the r package monitor,"Bioacoustics, vol. 25,no. 2, pp. 177–196, 2016.

[13] L. Neal, F. Briggs, R. Raich, and X. Z. Fern, "Time-frequency seg-mentation of bird song in noisy acoustic environments," in2011 IEEEInternational Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2011, pp. 2012–2015.

[14] M. Towsey, B. Planitz, A. Nantes, J. Wimmer, and P. Roe, "A toolboxfor animal call recognition,"Bioacoustics, vol. 21, no. 2, pp. 107–125,2012.

[15] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H.Frommolt, "Detecting bird sounds in a complex acoustic environmentand application to bioacoustic monitoring,"Pattern Recognition Letters,vol. 31, no. 12, pp. 1524–1534, 2010.

[16] J. Xie, J. G. Colonna, and J. Zhang, "Bioacoustic signal denoising: areview,"Artificial Intelligence Review, pp. 1–23, 2020.

[17] A. L. Borker, M. W. McKown, J. T. Ackerman, C. A. EAGLES-SMITH,B. R. Tershy, and D. A. Croll, "Vocal activity as a low cost and scalableindex of seabird colony size,"Conservation biology, vol. 28, no. 4, pp.1100–1108, 2014.

[18] D. Stowell and M. D. Plumbley, "Automatic large-scale classificationof bird sounds is strongly improved by unsupervised feature learning,"PeerJ, vol. 2, p. e488, 2014.

[19] E. Ozanich, A. Thode, P. Gerstoft, L. A. Freeman, and S. Freeman,"Unsupervised clustering of coral reef bioacoustics,"arXiv preprintarXiv:2012.09982, 2020.
F. Merchan, G. Echevers, H. Poveda, J. E. Sanchez-Galan, and H. M.Guzman, "Detection and identification of manatee individual vocal-izations in panamanian wetlands using spectrogram clustering,"TheJournal of the Acoustical Society of America, vol. 146, no. 3, pp. 1745–1757, 2019.

[20] P. Jancovic and M. K¨ok¨uer, "Bird species recognition using unsupervisedmodeling of individual vocalization elements,"IEEE/ACM Transactionson Audio, Speech, and Language Processing, vol. 27, no. 5, pp. 932–947, 2019.

[21] Xeno-canto. [Online]. Available: https://www.xeno-canto.org/

[22] S. Boll, “Suppression of acoustic noise in speech using spectral subtrac-tion,”IEEE Transactions on acoustics, speech, and signal processing,vol. 27, no. 2, pp. 113–120, 1979.
[23] J. Lim and A. Oppenheim, “All-pole modeling of degraded speech,”IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26,no. 3, pp. 197–210, 1978.
[24] B. M. Gur and C. Niezrecki, “Autocorrelation based denoising of man-atee vocalizations using the undecimated discrete wavelet transform,”The Journal of the Acoustical Society of America, vol. 122, no. 1, pp.188–199, 2007.
[25] N. Priyadarshani, S. Marsland, I. Castro, and A. Punchihewa, “Birdsongdenoising using wavelets,”PloS one, vol. 11, no. 1, p. e0146790, 2016