# Identificación Automática de Especies de Aves a través de su Canto

Lucas Genzelis, Emmanuel Rojas Fredini y Cristian Yones
*Trabajo Práctico final de "Procesamiento Digital de Señales", II-FICH-UNL.*

***Resumen*—El presente trabajo ha sido motivado por la necesidad de desarrollar un método que permita identificar la especie a la que pertenece un ave determinado a partir de su canto. Partiendo de la base de estudios previos que lograron demostrar la factibilidad de alcanzar este objetivo, se ha desarrollado un enfoque novedoso a utilizar en este proceso. El mismo compete, fundamentalmente, a la definición de las características a ser extraídas de las notas que componen el canto del ave. Debido a la complejidad propia de dicho canto, se ha optado por extraer una sucesión de medidas estadísticas tomadas a partir de cada nota. Dichas medidas son luego comparadas con medidas almacenadas en una base de datos, logrando de esta forma la identificación del ave en cuestión. La comparación se ha realizado utilizando dos métricas diferentes a fin de contrastar su desempeño.**

**Los resultados obtenidos mediante la aplicación de este procedimiento han sido favorables, demostrando su utilidad, y alcanzando el objetivo propuesto, identificando además aquella métrica que es conveniente utilizar en el proceso.**

***Palabras clave*—reconocimiento, identificación, aves, cantos**

## I. Introducción

Existe en el territorio nacional una gran diversidad de especies de aves. La mayoría de éstas pueden emitir sucesiones de sonidos elaborados, comúnmente denominadas cantos. El presente trabajo se plantea el objetivo de desarrollar un método que permita identificar la especie a la que pertenece un determinado ave, a través de dicho canto.

Los sonidos producidos por las aves presentan una estructura muy compleja. Su canto puede dividirse en cuatro niveles jerárquicos: notas, sílabas, frases y canción. La velocidad con que se producen cambios en las frecuencias presentes en estos sonidos es varios órdenes de magnitud más alta que en el ser humano, y su espectro mucho más amplio que el de éste.

Con el propósito de ilustrar lo dicho anteriormente, a continuación se muestran los espectrogramas correspondientes a porciones de canto de dos especies de aves diferentes:

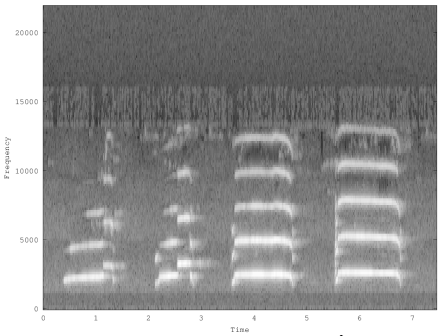


Fig. 1: Espectrograma del canto de un Águila Poma

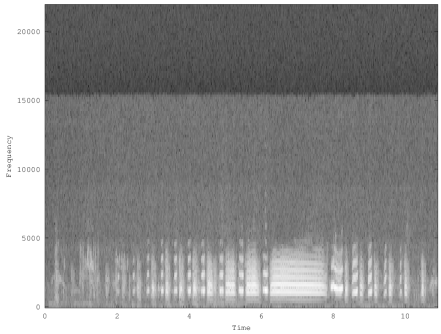


Fig. 2: Espectrograma del canto de un Pingüino

El método que se pretende desarrollar podría utilizarse para una diversidad de fines ecológicos, como en la realización de relevamiento de especies presentes en un determinado ecosistema, o en estudios de impacto ambiental sobre alteraciones realizadas sobre dichos ecosistemas (deforestaciones, construcción de caminos, etc.). Además existe posibilidad de explotación comercial de dicho método, en su utilización con fines recreacionales (por ejemplo, por los observadores de aves).

Sin embargo, relativamente poco ha sido hecho en este campo. Estudios anteriores han logrado demostrar la factibilidad del reconocimiento de especies de aves basado en su canto. Se han utilizado modelos ocultos de Markov, adaptados de los estudios de reconocimiento del habla, para modelar las probabilidades de transición entre sílabas del canto de un ave. También se ha intentado aproximar cada sílaba del canto de un ave por una onda sinusoidal modulada en amplitud y frecuencia, y comparar los parámetros de dichas ondas sinusoidales con parámetros pre-calculados para cada especie [1],[2],[3], [4].

## II. Materiales

El principal recurso necesario para la realización del presente trabajo son grabaciones de cantos de aves de diferentes especies, de forma que pueda ponerse a prueba la metodología desarrollada.

La mayor parte de estas grabaciones han sido obtenidas del DVD "Sonidos de Aves de Argentina y Uruguay", de Bernabe Lopez-Lanus.

Además se han obtenido grabaciones adicionales de diferentes fuentes de la Web, y en particular se obtuvo una grabación de una paloma de ciudad por intermedio del Prof. L. Di Persia.

## III. Metodología

La metodología aplicada se divide en dos procedimientos, referidos a la preparación de la base de datos que será utilizada durante la identificación, y a la identificación propiamente dicha.

El primero consiste en realizar la *separación en notas*, luego la *extracción de características* y por último la *generación de plantillas*, utilizando para esto *medidas de comparación*.

El segundo consiste en realizar la *separación en notas*, luego la *extracción de características* y por último la *clasificación*, utilizando para esto *medidas de comparación*.

Cada uno de estos puntos será explicado a continuación.

### *A. Separación en notas*

Es el proceso inicial, que consiste en dividir el canto del ave en unidades fundamentales cuyas características serán extraídas posteriormente. Para esto se toman ventanas solapadas de una longitud determinada y se aplican sobre el vector de audio $x$. Se calcula un mínimo de energía como

$$Emin = 0.5\ lwin\ \frac{\sum_{n=0}^{N-1}(x[n])^2}{N} \quad (1)$$

donde *lwin* es el número de muestras de cada ventana, y $N$ el número total de muestras.

El criterio tenido en cuenta para determinar cada unidad es:

-Identificar aquella ventana con una energía superior o igual a $1.2\ Emin$

-Identificar la siguiente ventana con una energía inferior a $0.8\ Emin$

-Si el número de muestras comprendido desde el inicio de la primer ventana hasta el fin de la segunda supera un mínimo establecido, se considera que entre estos límites se encuentra una unidad

-Se repite el procedimiento hasta haber recorrido todas las muestras de $x$

### *B. Extracción de características*

Previamente se ha hablado sobre la complejidad del canto de las aves y sobre la gran velocidad con que se producen cambios en las frecuencias presentes en los mismos. Estas características traen como consecuencia que los métodos normalmente aplicados al reconocimiento del habla humana no sean aplicables aquí. Por ejemplo, la cantidad de frecuencias presentes en cualquier porción de una unidad fundamental es tan alta que la noción de "frecuencias formantes" carece de utilidad práctica.

Por lo expuesto anteriormente, se ha optado por extraer, para cada unidad fundamental, una sucesión de medidas estadísticas representativas.

Para esto, se aplican sobre cada unidad fundamental ventanas solapadas (de Hamming, para suavizar las discontinuidades), y para cada una de ellas se procede de la siguiente forma:

-Se calcula su Transformada Discreta de Fourier $Xf$

-Se toman de $Xf$ los $n$ valores correspondientes a las frecuencias inferiores a la mitad de la frecuencia de muestreo

-A partir de dichos valores se calculan las siguientes medidas.

$$Media = \frac{\sum_{k=0}^{n-1} f_k\,|Xf_k|}{\sum_{k=0}^{n-1}|Xf_k|} \quad (2)$$

$$Varianza = \frac{\sum_{k=0}^{n-1}(f_k - Media)^2\,|Xf_k|}{\sum_{k=0}^{n-1}|Xf_k|} \quad (3)$$

$$Asimetría = \frac{\sum_{k=0}^{n-1}(f_k - Media)^3\,|Xf_k|}{\sum_{k=0}^{n-1}|Xf_k|} \quad (4)$$

Estas medidas estadísticas son organizadas en tres vectores, conteniendo éstos las secuencias de medias, varianzas y asimetrías, respectivamente.

### *C. Medidas de comparación*

Tanto para la generación de plantillas como para la clasificación del canto de un ave, es necesario disponer de una métrica que permita evaluar numéricamente la diferencia entre las características extraídas de dos unidades fundamentales.

Se implementaron dos métricas diferentes a fin de comparar los resultados obtenidos a través de ellas. Éstas son *Medida de la Distancia Normalizada*, y *Dynamic Time Warping* [5].

Luego, al comparar dos unidades fundamentales se aplicarán estas métricas evaluando la similaridad entre los vectores de medias, varianzas, y asimetrías, y se sumarán los valores obtenidos.

1. Medida de la distancia normalizada

   Se plantea la necesidad de definir una métrica normalizada que mida la semejanza entre dos vectores de características. No puede utilizarse el producto interno normalizado puesto que éste sólo tiene en cuenta el ángulo entre los vectores, y aquí es igualmente relevante la diferencia entre sus longitudes.

   Por lo tanto se decide utilizar la definición de la distancia Euclídea entre los vectores. Para normalizarla se la divide por la suma de los módulos de ambos vectores, obteniéndose así valores entre 0 y 1.

   $$D = \frac{norm(v1 - v2)}{norm(v1) + norm(v2)} \quad (5)$$

   donde el operador *norm* representa la norma L2.

2. Dynamic Time Warping

   Esta métrica permite evaluar numéricamente la similaridad entre dos vectores, teniendo en cuenta que la velocidad de las secuencias que representan podría no ser la misma, e incluso variar en el tiempo.

   El valor representativo de la similaridad entre los vectores es entonces el error persistente en el caso de coincidencia óptima entre ellos.

### *D. Generación de plantillas*

Antes de proceder a realizar la identificación de un ave, debe disponerse de una base de datos con información almacenada sobre diferentes especies. Esta etapa es equivalente a la de entrenamiento utilizada en las redes neuronales.

El procedimiento realizado para generar las plantillas es el siguiente:

-Para cada ave se lee un archivo de sonido con sus cantos, y éste es almacenado en un vector.

-Dicho vector es dividido en notas, según el procedimiento detallado previamente.

-Para cada nota identificada, se extraen sus características.

-Se almacenan las características de la primer nota.

-Para cada una de las notas siguientes, se utiliza una de las medidas de comparación definidas para evaluar la similaridad entre la nota en cuestión, y las notas que ya han sido almacenadas. Si en todos los casos la diferencia es mayor que un cierto umbral, se almacenan las características de la nueva nota. Si en algún caso esta diferencia es menor que el umbral, se considera que se trata de la misma nota, y se promedian sus características con las de la nota almacenada. Junto a cada nota se almacena el número de veces que se ha encontrado dicha nota, de manera que este valor pueda ser utilizado en el promediado para conservar la uniformidad de los pesos otorgados a las notas que han sido promediadas. Es decir, sea *caract* el vector de características de una nota, *N* el número de veces que la correspondiente nota ha sido identificada, y $caract_{n+1}$ el vector de características de una nueva ocurrencia de dicha nota, se toma:

$$caract = \frac{caract \,.\, N + caract_{n+1}}{N+1} \tag{6}$$

### *E. Clasificación*

Este punto corresponde a la implementación del propósito del presente trabajo, es decir, a la identificación de la especie a la que pertenece un determinado ave a partir de su canto.

Para esto se procede de la siguiente forma:

-Se lee un archivo de sonido con el canto del ave, y se lo divide en notas.

-Para cada nota, se extraen sus vectores de características.

-Utilizando una de las medidas de comparación, se evalúa la diferencia entre cada nota y cada especie de ave de la que se hayan almacenado plantillas. Para medir la diferencia entre una nota y un ave, se calcula la diferencia entre la nota en cuestión y todas las notas que se hayan almacenado para ese ave, y se toma la menor de todas estas diferencias.

-Para cada ave, se suman las diferencias de cada nota a dicho ave.

-La menor de todas estas diferencias permitirá identificar a qué ave corresponde el canto del archivo de audio.

## IV. Análisis de Resultados

Se realizaron pruebas utilizando particiones disjuntas de entrenamiento y de prueba durante la realización del entrenamiento y del reconocimiento, respectivamente.

A continuación se analizarán los resultados obtenidos mediante la aplicación de cada una de las métricas desarrolladas en la sección previa.

### *A. Medida de la distancia normalizada*

En la Tabla I puede observarse, en la primera columna, la lista de especies sobre las que se realizaron las pruebas, en la segunda columna, la duración de la muestra de entrenamiento, y en la tercer columna, la duración de la muestra de prueba. En la Tabla II pueden verse los resultados del proceso de identificación usando las particiones antes mencionadas. En la tercer columna figura un coeficiente que se ha definido como métrica de la seguridad de la identificación.

Dicho coeficiente puede definirse de diversas formas, siempre de manera que se encuentre normalizado entre 0 y 1. Por ejemplo, sea *diff* el vector de diferencias totales de cada ave con respecto al canto analizado, *N* el número total de aves, y $diff_{min}$ el valor mínimo de este vector, dicho coeficiente podría definirse como:

$$\alpha = 1 - \frac{diff_{min}}{\sum_{i=0}^{N-1} diff[i]} \tag{7}$$

Otra definición podría darse utilizando los dos valores mínimos del vector de diferencias $diff_1$ y $diff_2$ ($diff_1 < diff_2$):

$$\alpha = \frac{diff_2 - diff_1}{diff_2} \tag{8}$$

Los datos presentados en la Tabla II corresponden a la definición dada en (7).

TABLA I

Duraciones de los archivos de entrenamiento y pruebas para distintas especies de aves

| Especie | Tiempo de entrenamiento [s] | Tiempo de prueba [s] |
|---|---|---|
| Águila Poma PD | 42 | 7 |
| Aguilucho Andino | 12 | 1 |
| Batara Negro | 16 | 2 |
| Chiflón | 21 | 5 |
| Chimango | 21 | 1 |
| Inambú Montaraz | 51 | 19 |
| Paloma | 103 | 28 |
| Pato Overo | 44 | 4 |
| Pingüino Patagónico | 9 | 13 |
| Tataupa Listado | 50 | 6 |

TABLA II

Resultados de las pruebas de identificación con su respectivo nivel de confianza

| Especie | Especie Reconocida | Confianza |
|---|---|---|
| Águila Poma PD | Águila Poma PD | 0,938399 |
| Aguilucho Andino | Aguilucho Andino | 0.942188 |
| Batara Negro | Batara Negro | 0.925721 |
| Chiflón | Chiflón | 0.954115 |
| Chimango | Chimango | 0.962602 |
| Inambú Montaraz | Inambú Montaraz | 0.932475 |
| Paloma | Paloma | 0.926347 |
| Pato Overo | Pato Overo | 0.945791 |
| Pingüino Patagónico | Pingüino Patagónico | 0.929857 |
| Tataupa Listado | Tataupa Listado | 0.930769 |

Debe destacarse que tanto las muestras de entrenamiento como las muestras de prueba que se han utilizado se hallan contaminadas con ruido, puesto que han sido tomadas en el ambiente natural; y no obstante esto, se logra un 100% de exactitud en el reconocimiento.

Por lo tanto, puede afirmarse que los resultados obtenidos han sido completamente satisfactorios.

### *B. Dynamic Time Warping*

No se obtuvieron resultados satisfactorios mediante la aplicación de esta métrica. En primer lugar, debido a su tiempo de ejecución cuadrático, que trae como consecuencia grandes esperas a la hora de realizar la identificación y, más aún, la generación de plantillas.

Además, en las pruebas realizadas con plantillas generadas a partir del canto de un número acotado de aves (donde deberían observarse mejores resultados, siempre que se intente identificar un ave presente en la base de datos), se obtuvieron muy pocas identificaciones correctas.

## V. Conclusiones y Trabajos Futuros

Como se ha demostrado, la fiabilidad del método propuesto es altamente dependiente de la medida de comparación utilizada.

Es claro que el método Dynamic Time Warping no proporciona una métrica adecuada para el problema en análisis.

Por otra parte, la medida de la distancia normalizada ha probado ser una métrica rápida y efectiva en la aplicación en estudio.

El hecho de que las muestras de entrenamiento hayan sido tomadas del ambiente natural, con el ruido propio de éste y de otras aves cercanas, pone en evidencia la robustez del método que se ha desarrollado.

En el futuro, podría utilizarse el presente trabajo para implementar un sistema de tiempo real que pueda embeberse en un dispositivo portátil, a fin de comercializarse o utilizarse con fines de estudios ambientales.

## Referencias

[1] E. Anderson, A. S. Dave, and D. Margoliash, “Template-based automatic recognition of birdsong syllables from continuous recordings,” J. Acoust. Soc. Am., vol. 100, pp. 1209–1219, Agosto 1996.

[2] J. A. Kogan and D. Margoliash, “Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study,” J. Acoust. Soc. Am., vol. 103, pp. 2185–2196,Abril 1998.

[3] A. L. McIlraith and H. C. Card, “Birdsong recognition using backpropagation and multivariate statistics,” IEEE Trans. Signal Processing, vol. 45, pp. 2740–2748, Noviembre 1997.

[4] A. Härmä, “Automatic identification of bird species based on sinusoidal modeling of syllables ,” IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.

[5] C. S. Myers, L. R. Rabiner, “A comparative study of several dynamic time-warping algorithms for connected word recognition.” The Bell System Technical Journal, vol. 60, pp. 1389-1409, Septiembre 1981.