07-007

# FEATURE ENGINEERING FOR SOUND SIGNAL CLASSIFICATION

*Gómez Bellido, Jesús [(1)]; Luque Sendra, Amalia [(1)]; Carrasco Muñoz, Alejandro [(1)]*

[(1)] Universidad de Sevilla

The classification of sound signals is an area in which there is a great research activity with very diverse applications in numerous areas of science and engineering. These applications include, among its first steps, the need to represent sound through a set of features. The set of techniques that allow the transfer of raw data to the most appropriate characteristics for its subsequent application is commonly known as features engineering.On the other hand, the processing and classification of anura songs have attracted the attention of the scientific community, both from a biological point of view, and as indicators of climate change. This work focuses on the parameters used to characterize the anura sounds. Jobs can be grouped into three categories: 1. Comparison of alternatives in the characterization of auro sounds. Determination of the optimal values of such characterization 2. Analysis of the role played by the symmetry of integral transformations in the characterization of sounds based on cepstral coefficients. 3. Analysis of the computational effort required in the different stages of the characterization and classification process of anura sounds.

*Keywords: feature engineering; classification; sound processing*

# INGENIERÍA DE CARACTERÍSTICAS PARA CLASIFICACIÓN DE SEÑALES SONORAS

La clasificación de señales sonoras es un área en el que existe una gran actividad investigadora con aplicaciones muy diversas en numerosas áreas de la ciencia y la ingeniería. Estas aplicaciones incluyen, entre sus primeros pasos, la necesidad de representar el sonido mediante un conjunto de características. Al conjunto de técnicas que permiten pasar de los datos en brutos a las características más adecuadas para su aplicación posterior se le conoce comúnmente como ingeniería de características. Por otro lado, el procesamiento y la clasificación de los cantos de anuros han atraído la atención de la comunidad científica, tanto desde un punto de vista biológico, cuanto como indicadores del cambio climático. Este trabajo se centra en los parámetros utilizados para caracterizar los cantos de anuros. Los trabajos se pueden agrupar en tres categorías: 1. Comparación de alternativas en la caracterización de cantos de anuros. Determinación de los valores óptimos de dicha caracterización 2. Análisis del papel que juega la simetría de las transformaciones integrales en la caracterización de sonidos basada en coeficientes cepstrales. 3. Análisis del esfuerzo computacional requerido en las distintas etapas del proceso de caracterización y clasificación de cantos de anuros.

*Palabras clave: ingeniería de características; clasificación; procesamiento de sonidos*

Correspondencia: Amalia Luque Sendra amalialuque@us.es

©2020 by the authors. Licensee AEIPRO, Spain. This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introducción

La clasificación de señales sonoras es un ámbito en el que existe una gran actividad investigadora con aplicaciones muy diversas en numerosas áreas de la ciencia y la ingeniería. Estas aplicaciones incluyen, entre sus primeros pasos, la necesidad de representar el sonido mediante un conjunto de características (denominadas también parámetros, variables, rasgos o features). Al conjunto de técnicas que permiten pasar de los datos en brutos a las características más adecuadas para su aplicación posterior se le conoce comúnmente como ingeniería de características.

Por otro lado, el procesamiento y la clasificación de los cantos de anuros (sonidos emitidos por estas especies animales como parte de su actividad sexual) han atraído la atención de la comunidad científica, tanto desde un punto de vista biológico, cuanto como indicadores del cambio climático.

**Figura 1: Extracción de características de cantos de anuros como indicador de cambio climático**


El presente trabajo recoge los trabajos de investigación realizados por los autores en el campo de la ingeniería de características para la mejor clasificación de sonidos correspondientes a diferentes cantos de anuros. El documento hunde sus raíces en trabajos previos (Larios et al, 2012; Luque, Larios, Personal., Barbancho, y León, 2016) para el despliegue de redes de sensores en el entorno de la monitorización medioambiental, así como el posterior tratamiento de los datos obtenidos. En relación al procesamiento de cantos de anuros, el punto de partida es el desarrollo de unos clasificadores simples basados en parámetros MPEG-7.

A partir de estos antecedentes se ponen en marcha dos líneas de trabajo. La primera línea caracteriza los cantos de anuros mediante parámetros MPEG-7 y trata de explorar distintos algoritmos de clasificación incluyendo técnicas no secuenciales, técnicas que consideran la secuencialidad de los sonidos y técnicas de clasificación de series de puntuaciones. Esta primera línea ha sido desarrollada en (Luque, Romero-Lemos, Carrasco, Barbancho, 2018; Luque, Romero-Lemos, Carrasco, González-Abril, 2018) y antecede lógica y cronológicamente a la investigación que se detalla en este documento.

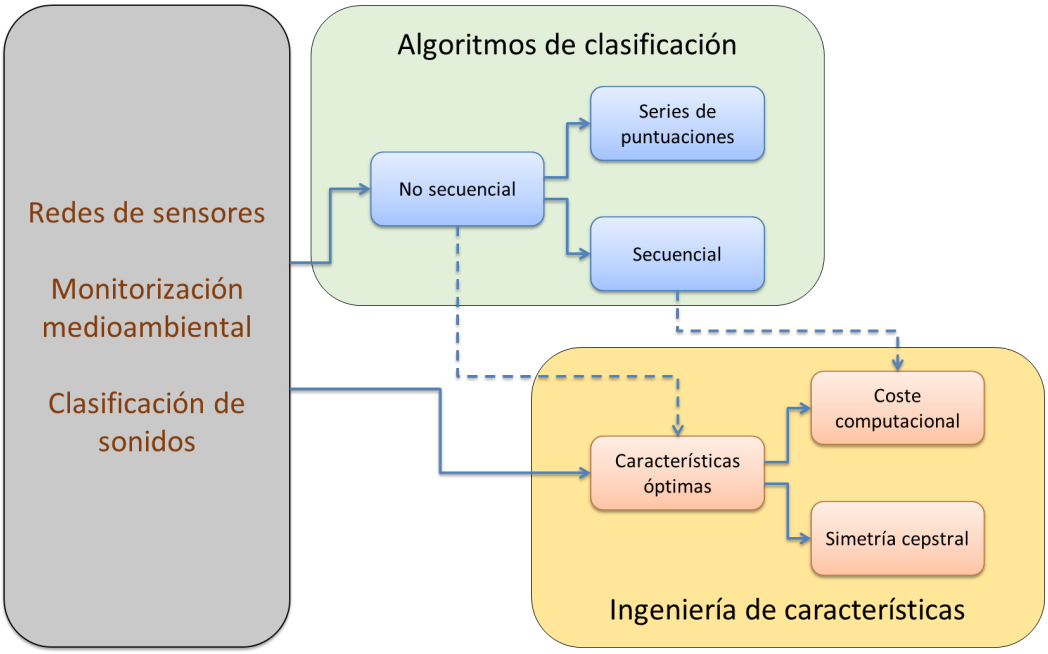
La segunda línea, que es el objeto del presente trabajo, se centra en los parámetros utilizados para caracterizar los cantos de anuros, sin considerar nuevas mejoras sobre los algoritmos de clasificación.

Los trabajos se pueden agrupar en tres categorías

- Comparación de alternativas en la caracterización de cantos de anuros. Determinación de los valores óptimos de dicha caracterización. Esta actividad constituye el núcleo central de la investigación realizada (Luque, Gómez-Bellido, Carrasco, y Barbancho, 2018)

- Análisis del papel que juega la simetría de las transformaciones integrales (transformada de Fourier y transformada coseno) en la caracterización de sonidos basada en coeficientes cepstrales (Luque, Gómez-Bellido, Carrasco, Barbancho, 2019).
- Análisis del esfuerzo computacional requerido en las distintas etapas del proceso de caracterización y clasificación de cantos de anuros (Luque, Gómez-Bellido, Carrasco, Personal y León, 2017).

**Figura 2: Categorías de trabajos**




## 2. Metodología

Para las pruebas se han utilizado sonidos reales de anuros proporcionados por el Museo Nacional de Ciencias Naturales (Gómez-García, Moro-Velázquez, Godino-Llorente, 2019.) (código de colección a partir de FZ0496). Los sonidos corresponden a las especies: epidalea calamita (sapo corredor) y alytes obstetricans (sapo partero común), con un total de 868 grabaciones que contienen 4 clases de sonidos:

1. Epidalea calamita; canto de apareamiento (369 grabaciones),

2. Epidalea calamita; canto en suelta (63 grabaciones),

3. Alytes obstetricans; canto de apareamiento (419 grabaciones),

4. Alytes obstetricans; canto de socorro (17 grabaciones).

En total, se ha analizado 4343 segundos (1 h 13 min) de grabaciones, con una duración media de 5 segundos por grabación. Los sonidos han sido grabados en cinco lugares diferentes (cuatro en España y uno en Portugal) usando un micrófono ME80 de Sennheiser (Wedemark, Alemania), tema discutido en detalle en (Luque, Larios, Personal., Barbancho, y León, 2016), y posteriormente se muestrean a 44,1 kHz. Una característica común de todas las grabaciones es que han sido tomadas en su hábitat natural, con un ruido ambiental muy significativo (viento, agua, lluvia, tráfico, voces, etc.), lo que supone un reto adicional en el proceso de clasificación. La distribución de la relación señal/ruido (SNR) para cada clase de sonido, aunque en algunas grabaciones tienen un valor mucho menor, el conjunto de datos

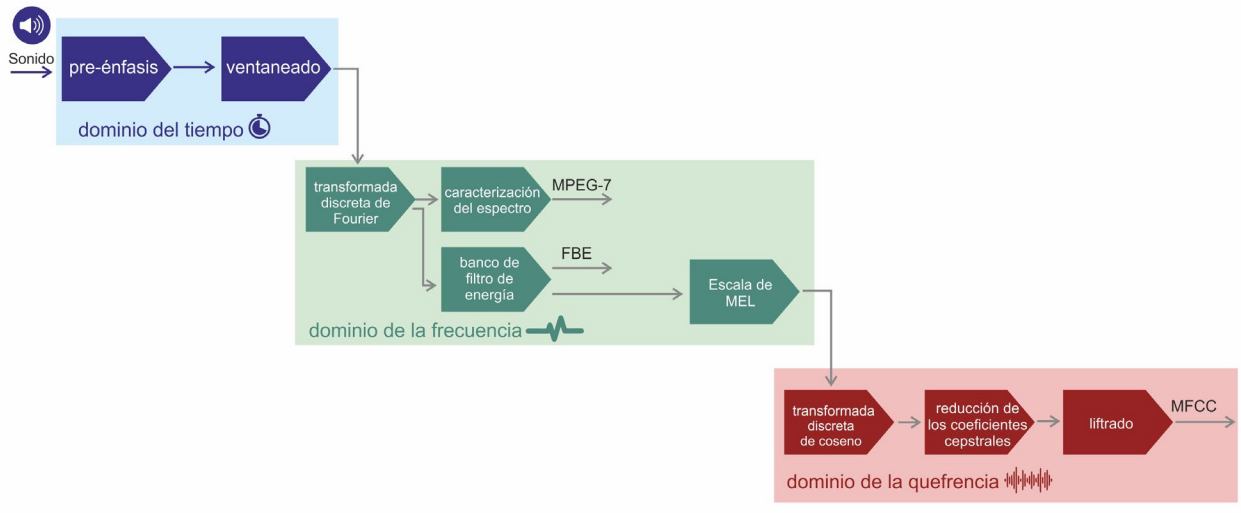
presenta un valor medio de SNR global de 35 dB. Para realizar una clasificación supervisada, ciertos sonidos deben ser seleccionados como patrones (para ser usados en la fase de entrenamiento) mientras que otros son empleados para las pruebas. Una práctica común es dividir el conjunto de datos en varios subconjuntos disociados y aplicar una técnica de validación cruzada.

Figura 3: Datos de los sonidos utilizados

| Clase | Sonidos | | Patrones | | |
|---|---|---|---|---|---|
| | Núm. grab. | Tiempo (s) | Núm. grab. | Tiempo del patrón (s) | Tiempo total (s) |
| Ep. cal. apareamiento | 369(43 %) | 1853 | 4 | 13, 89 | 20, 39 |
| Ep. cal. suelta | 63(7 %) | 311 | 3 | 0,99 | 14,56 |
| Al. ob. apareamiento | 419(48 %) | 2, 096 | 4 | 1,09 | 19,72 |
| Al. ob. peligro | 17(2 %) | 83 | 2 | 3,30 | 9,80 |
| Sil./Ruido | - | - | - | 45, 20 | - |
| Total | 868 | 4, 343 | 13 | 64,47 | 64,47 |

Sin embargo, el uso de estas grabaciones ruidosas como patrones puede llevar a una disminución en el rendimiento de la clasificación. De ahí que surjan otros enfoques como alternativa a la validación cruzada. En nuestro caso, las grabaciones con ruido de fondo relativamente bajo, que fueron cuidadosamente seleccionadas por biólogos e ingenieros de sonido, son las que se han utilizado como patrones. Este enfoque, generalmente llamado selección de instancia o ejemplo, se recomienda para aumentar el ritmo de aprendizaje centrando la atención en ejemplos informativos (Gonzalez-Abril, Haydemar Nuñez, and Velasco, 2014).

**Figura 4: Metodología**



El primer paso para representar un sonido es dividirlo en frames de duración fija.

La representación del espectro suele basarse en los frames obtenidos en el paso anterior. El procedimiento para obtener un vector de valores que represente un frame se denomina extracción de características

Sin embargo, estas características no consideran la característica secuencial intrínseca de la evolución temporal del sonido. Por lo tanto, esta información secuencial debe ser añadida construyendo nuevas características.

Para abordar el proceso de clasificación, el conjunto de datos de sonido tiene que ser dividido en 3 subconjuntos. En primer lugar, se han utilizado como patrones las grabaciones con un ruido de fondo relativamente bajo, que fueron seleccionadas cuidadosamente por biólogos e ingenieros de sonido.

La definición de las métricas de desempeño de clasificación más adecuadas representa un aspecto clave en la evaluación de los procedimientos (Sturm, 2014). Con el fin de comparar los resultados obtenidos para cada clasificador y cada combinación de características, se pueden definir varias métricas de desempeño (Sokolova y Lapalme, 2009.), todas ellas basadas en la matriz de confusión binaria.

Se han propuesto diferentes implementaciones o alternativas para el análisis de los cantos de anuros. Sin embargo, desde el punto de vista de la implementación, estos algoritmos no son triviales y pueden requerir mucho tiempo de ejecución. En este sentido, un análisis exhaustivo del tiempo de cada etapa es esencial para garantizar la aplicación en tiempo real.

## 3. Resultados

En los trabajos de investigación realizados por los autores se ha estudiado la mejor forma de representar un sonido aplicando estos resultados a la clasificación de cantos de anuros. En concreto los temas abordados han sido los siguientes:

Se ha realizado un resumen del estado del arte, en el que se presenta la aplicación de clasificación de sonidos de anuros y se justifica como indicador de cambio climático. Se describen y referencian las principales técnicas de caracterización de sonidos. Se presenta un conjunto de 18 parámetros basados en la norma MPEG-7 utilizados en trabajos previos. Se detalla la técnica de obtención de los MFCC. Se discute el uso de clasificadores CNN y RNN en base a sonidos “en bruto” (sin necesidad de extracción de características) y se justifica su no utilización. Se presentan diversos clasificadores habitualmente utilizados en el ámbito del aprendizaje automático *(machine learning*).

**Figura 5: Resultados. Elaboración propia adaptada de (Pérez, 2018 y Tang, 2019)**

Se comparan el uso de características MPEG-7 y MFCC en la representación de sonidos para su posterior clasificación. Se explora la influencia de los diversos parámetros que intervienen en la extracción de los MFCC. Se selecciona y ajusta el procedimiento de extracción de MFCC para una clasificación óptima. Se analiza la mejor forma de obtener coeficientes cepstrales y se justifica en base a la simetría de las transformaciones integrales (transformada de Fourier y transformada coseno). Se realiza un análisis del esfuerzo computacional requerido en las distintas etapas del proceso de caracterización y clasificación de cantos de anuros. Se aplican estas propuestas a la clasificación de un conjunto de 868 grabaciones reales realizadas en campo, con más de hora y media de duración acumulada.

## 4. Discusión y conclusiones

La clasificación de sonidos mejora sensiblemente (7 puntos de la métrica GM) si se usan MFCC con respecto al uso de parámetros MPEG-7. La optimización de los parámetros de extracción de los MFCC mejora la clasificación (2 puntos de la métrica GM) La extracción de los MFCC puede ser descompuesta en distintas etapas cada una de las cuales ofrece una mejora diferencial en los resultados de clasificación con respecto al uso de parámetros MPEG-7 que se puede desglosar de la siguiente forma:

**Tabla 1. Principales resultados obtenidos**

| | GM DIFERENCIAL | GM ACUMULADA |
|---|---|---|
| MPEG-7 (valores base: baseline) | 0 | 0 |
| Energía de banco de filtros (FBE: Filter Bank Energy) | -2 | -2 |
| FBE in log-scale | +2 | 0 |
| FBE in mel-log-scale | +5 | +5 |
| FBE in mel-log-scale (optimum options) | +1.5 | +6.5 |
| MFCC (DCT of the FBE in mel-log-scale) | -1.5 | +5 |
| MFCC with optimum frame duration | +0.5 | +5.5 |
| MFCC with optimum options | +1.5 | +7 |

Este resultado puede resumirse en las siguientes conclusiones

- El uso de escalas logarítmicas mejora 2 puntos.
- El uso de la escala mel mejora 5 puntos.
- El paso al dominio cepstral (DCT del FBE en escala mel logarítmica), es decir, el uso de los MFCC con parámetros de extracción estándar, no supone una mejora de los resultados de clasificación.
- La optimización de los parámetros de extracción de los MFCC mejora 2 puntos.

    - La optimización de los parámetros de extracción de los FBE en escala mel logarítmica mejora 1.5 puntos.
    - Cuando se usan parámetros de extracción óptimos, el paso al dominio cepstral mejora 0.5 puntos.
- El uso de un menor número de coeficientes en la caracterización del sonido reduce las prestaciones en un valor aproximado de -0.3 puntos por coeficiente.
    - Cuando el número de coeficientes es muy reducido, los MFCC presentan una clara ventaja sobre los FBE en escala mel logarítmica.
- En el proceso de extracción de los MFCC el paso del dominio espectral al dominio cepstral se puede realizar utilizando la transformada de Fourier (DFT) o la transformada coseno (DCT).
    - El uso de la DCT introduce un 30% menos de error que si se utiliza la DFT.
    - Los coeficientes obtenidos mediante DCT están sensiblemente menos correlados que los obtenidos mediante DFT. Esta menor correlación producirá mejores resultados de clasificación.
    - Las mejoras de la DCT sobre la DFT se deben a la simetría del espectro.
    - Estos mejores resultados de la DCT pueden extrapolarse a otros sonidos y no exclusivamente a los cantos de anuros
- En relación al coste computacional del proceso de clasificación de sonidos, las conclusiones pueden desglosarse en los siguientes términos:
    - El coste de la extracción de los parámetros MFCC es mucho menor que el de los parámetros MPEG-7 (por un factor de 60).
    - El coste de la construcción de parámetros es relativamente bajo y del mismo orden de magnitud que el de extracción de parámetros MFCC.
    - El coste del algoritmo de clasificación depende del clasificador utilizado aunque se puede afirmar que, en general, es del mismo orden de magnitud que el de extracción de parámetros MFCC.
    - El proceso global de clasificación puede realizarse en una fracción (pequeña) de la duración de un frame de sonido, es decir, que puede ejecutarse en tiempo real.

Derivadas de este trabajo y como continuación de la investigación realizada, se plantean distintas líneas e ideas que permitirían continuar el desarrollo iniciado:

- Extensión de los resultados de esta tesis a un conjunto más amplio de sonidos, incrementando tanto el tamaño global como el tipo de sonido.
- Estudio de representación en bruto de los sonidos como entrada de clasificadores avanzados (por ejemplo CNN y RNN).
- Desarrollo de técnicas de reducción de dimensionalidad (feature selection) para disminución del número de características necesarias en la representación de sonidos.
- Implementación de los procesos de representación y clasificación de sonidos en procesadores de bajo coste y bajo consumo.
- Estudio de métricas de clasificación con varias clases desequilibradas en las que se valoren adecuadamente el compromiso en el resultado de cada clase

- Se presentan diversos clasificadores habitualmente utilizados en el ámbito del aprendizaje automático (machine learning).

## Agradecimientos

Los autores agradecen a Ana de las Heras García de Vinuesa su ayuda en la elaboración de las imágenes y en la revisión final de formato.

## 6. Referencias

Gómez-García, J.A., Moro-Velázquez L., and Godino Llorente, J.I.. On the design of automatic voice condition analysis systems. part ii: Review of speaker recognition techniques and study on the effects of different variability factors. Biomedical Signal Processing and Control, 48:128–143, 2019.

Gonzalez-Abril, L, Haydemar Nuñez, C., and Velasco, F.. Gsvm: An svm for handling imbalanced accuracy between classes inbi-classification problems. Applied Soft Computing, 17:23–31, 2014.

Larios, D. F., Barbancho, J., Rodríguez, G., Sevillano, J. L., Molina, F. J., & León, C. (2012). Energy efficient wireless sensor network communications based on computational intelligent data fusion for environmental monitoring. IET communications, 6(14), 2189-2197.

Luque, A., Gómez-Bellido, J., Carrasco, A., & Barbancho, J. (2019). Exploiting the symmetry of integral transforms for featuring anuran calls. Symmetry, 11(3), 405.

Luque, A., Gómez-Bellido, J., Carrasco, A., & Barbancho, J. (2018). Optimal Representation of Anuran Call Spectrum in Environmental Monitoring Systems Using Wireless Sensor Networks. Sensors, 18(6), 1803.

Luque, A., Romero-Lemos, J., Carrasco, A., & Barbancho, J. (2018). Non-sequential automatic classification of anuran sounds for the estimation of climate-change indicators. Expert Systems with Applications, 95, 248-260.

Luque, A., Romero-Lemos, J., Carrasco, A., & Barbancho, J. (2018). Improving Classification Algorithms by Considering Score Series in Wireless Acoustic Sensor Networks. Sensors, 18(8), 2465.

Luque, A., Romero-Lemos, J., Carrasco, A., & Gonzalez-Abril, L. (2018). Temporally-aware algorithms for the classification of anuran sounds. PeerJ, 6, e4732.

Luque, A., Gómez-Bellido, J., Carrasco, A., Personal, E., & Leon, C. (2017). Evaluation of the processing times in anuran sound classification. Wireless Communications and Mobile Computing, 2017.

Luque, A., Gómez-Bellido, J., Carrasco, A., Falomir, Z., & Abril, L. G. (2017, October). MFCC vs. MPEG-7 Frame Description for Anuran Sound Classification. In CCIA (pp. 206-214).

Luque, J., Larios, D. F., Personal, E., Barbancho, J., & León, C. (2016). Evaluation of MPEG-7-based audio descriptors for animal voice recognition over wireless acoustic sensor networks. Sensors, 16(5), 717.

Pérez, W (2018). Mendely y estado del arte. Obtenido de https://elsancarlistau.com/2018/03/07/mendely-y-estado-del-arte/

Sturm, B. A simple method to determine if a music information retrieval system is a “horse”. IEEE Transactions on Multimedia, 16(6):1636–1644, 2014.

Sokolova, M, Lapalme, G. A systematic analysis of performance measures for classification tasks. Information processing & management,45(4):427–437, 2009.

Tang, B (2019) Data Collection and Feature Extraction for Machine Learning. Obtenido de https://medium.com/ai³-theory-practice-business/data-collection-and-feature-extraction-for-machine-learning-98f976401378

## Comunicación alineada con los Objetivos de Desarrollo Sostenible