

# **ASPECTOS DE LA SIMETRÍA DE LAS TRANSFORMACIONES INTEGRALES EN LA CARACTERIZACIÓN DE SONIDOS**

**Jesús Gómez Bellido<sup>1,\*</sup>, Amalia Luque Sendra<sup>1</sup>, Alejandro Carrasco Muñoz<sup>2</sup>, Julio Barbancho Concejero<sup>2</sup>**

<sup>1</sup> *Departamento de Ingeniería del Diseño. Escuela Politécnica Superior. Universidad de Sevilla.*

<sup>2</sup> *Departamento de Tecnología Electrónica. Escuela Técnica Superior de Ingeniería Informática. Universidad de Sevilla.*

E-mail de correspondencia: [jesgombel@outlook.es](mailto:jesgombel@outlook.es)

## **Resumen**

La aplicación de técnicas de aprendizaje automático a señales de sonido requiere la caracterización previa de dichas señales. A menudo, la descripción del espectro sonoro se realiza utilizando coeficientes cepstrales. En este trabajo, se compara el desempeño en la obtención de coeficientes cepstrales mediante dos transformaciones integrales, la Transformación discreta de Fourier (DFT) y la Transformación discreta de coseno (DCT). Se puede demostrar que DCT supera a DFT en la tarea de representar los espectros de sonido. Esta mejora se debe a la simetría del espectro y no a ninguna ventaja intrínseca de DCT. Además, las características de MFCC obtenidas con DCT están notablemente menos correlacionadas que las obtenidas con DFT, lo que hará que las funciones MFCC basadas en DCT sean más potentes en algoritmos de clasificación posteriores.

## **1. Introducción**

El procesamiento automático de las señales de sonido es un tema muy activo en muchos campos de la ciencia y la ingeniería que tiene aplicaciones en múltiples áreas. Estas aplicaciones suelen incluir, entre sus primeros pasos, la caracterización del sonido: un proceso que se conoce comúnmente como extracción de características.

Los Coeficientes Cepstrales de Frecuencia Mel (MFCC) son probablemente el conjunto de características más ampliamente utilizados en la caracterización del sonido y la mayoría de las aplicaciones de procesamiento de sonido mencionadas anteriormente se basan en su uso.

En trabajos previos del grupo de investigación, se han propuesto varios clasificadores para clasificación de sonidos que utilizan procedimientos no secuenciales (Luque *et al.*, 2018b) o algoritmos con conciencia temporal (Luque *et al.*, 2018a). En Luque *et al.* (2017) la comparación de MPEG-7 y MFCC se realiza tanto en términos de rendimiento de clasificación como de coste computacional.

Finalmente, los valores óptimos de las opciones de MFCC para la clasificación de nuestro caso de estudio, los cantos de anuros se derivan en Luque *et al.* (2018).

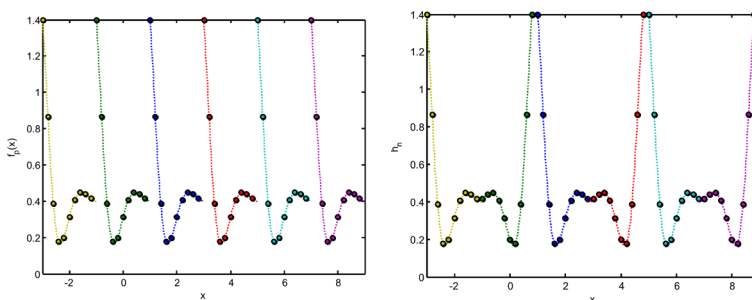
## 2. Materiales y métodos

El proceso de extracción de las características MFCC de las  $n$  muestras de un determinado sonido requiere 7 pasos en 3 dominios diferentes (tiempo, frecuencia, cepstral).

En este proceso, las transformaciones integrales se utilizan dos veces: en primer lugar, para pasar del dominio del tiempo al dominio de la frecuencia, haciendo uso de la DFT; y finalmente para pasar al dominio cepstral.

En este estudio, se va a analizar las propiedades simétricas de la DFT de la primera transformación para encontrar la selección de la transformación integral más apropiada para el último cambio de dominio.

Por la propia definición de la DFT, para representar el espectro de un fragmento de una señal de sonido se supone una repetición periódica del fragmento a analizar (Luque *et al.*, 2019), como se representa en la Figura 1 (panel A). Mientras que DCT considera una repetición periódica de este fragmento y su simétrica (Luque *et al.*, 2019), como se representa en la Figura 1 (panel B).



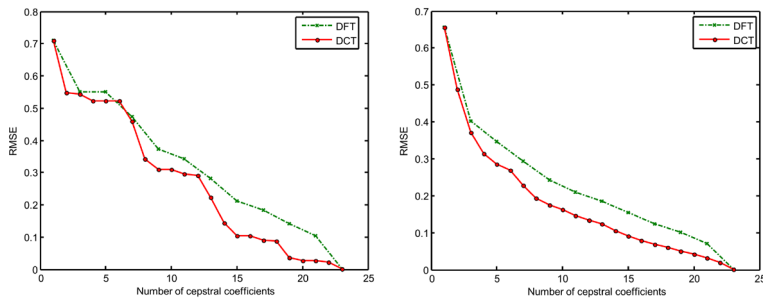
**Figura 1.** A) Representación periódica de un fragmento conocido. B) Representación periódica de un fragmento conocido y su simetría.

### 3. Resultados y discusión

En la Figura 2, podemos ver el valor de la raíz del error cuadrático medio en función del número de coeficientes cepstrales utilizados para la recuperación del espectro para cada una de las transformadas

En la figura de la izquierda, sería para un sólo *frame* elegido aleatoriamente, mientras que a la derecha vemos la representación del error para todo el conjunto de datos.

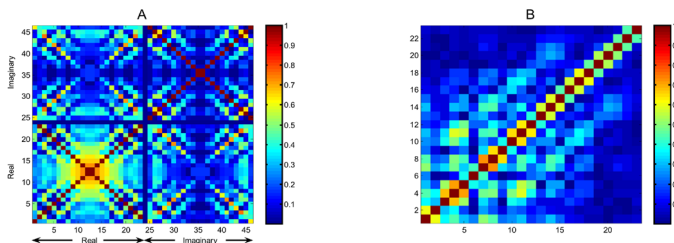
Como se puede ver, el error por norma general es significativamente menor en el caso de la DCT.



**Figura 2.** Error cuadrático medio empleando DFT y DCT utilizando un número  $C$  de coeficientes cepstrales a) para un *frame*, b) para todo el conjunto de dato.

Otra de las características de los clasificadores es que obtienen mejores resultados si sus características de entrada están relacionadas con un bajo nivel de correlación. Puesto que un algoritmo de clasificación que incluye una nueva característica altamente correlacionada con características anteriores no aporta nueva información significativa y, por lo tanto, casi no se espera mejora en la clasificación.

En la Figura 3, vemos en la imagen de la izquierda las correlaciones correspondientes a la DFT, mientras que en la imagen derecha se muestran las correspondientes a la DCT.



**Figura 3.** Correlación entre los coeficientes cepstrales que describen los espectros de mel-FBE para DFT (izquierda) y DCT (derecha).

Simplemente considerando los colores en estas imágenes, está claro que los coeficientes DCT están menos correlacionados.

#### 4. Conclusiones

- La Transformada Coseno Discreta (DCT) supera a la Transformada Discreta de Fourier (DFT) en la tarea de representar espectros de sonido. Esta mejora se debe a la simetría del espectro y no a ninguna ventaja intrínseca de la DCT.
- Las características MFCC obtenidas mediante la DCT están notablemente menos correladas que las obtenidas mediante la DFT. Este resultado hará que las características MFCC basadas en Transformada Coseno Discreta sean más potentes en algoritmos de clasificación posteriores.
- Puede comprobarse como la DCT tiene un tiempo de computación mayor que la DFT, pero que no afecta a la clasificación en tiempo real.

#### Agradecimientos

A los grupos de investigación TEP022 y TIC150 de la Universidad de Sevilla por los conocimientos adquiridos y los trabajos previos realizados.

A la Escuela Politécnica Superior de Sevilla por la organización de la Jornadas Doctorales que ha permitido la publicación del presente trabajo.

#### Referencias bibliográficas

- Luque, A., Gómez-Bellido, J., Carrasco, A., Personal, E., y León, C.** (2017). Evaluation of the processing times in anuran sound classification. *Wireless Communications and Mobile Computing*, 2017. Article ID 8079846. <https://doi.org/10.1155/2017/8079846>
- Luque, A., Gómez-Bellido, J., Carrasco, A., y Barbancho, J.** (2018). Optimal Representation of Anuran Call Spectrum in Environmental Monitoring Systems Using Wireless Sensor Networks. *Sensors*, 18(6), 1803. <https://doi.org/10.3390/s18061803>
- Luque, A., Gómez-Bellido, J., Carrasco, A., y Barbancho, J.** (2019). Exploiting the Symmetry of Integral Transforms for Featuring Anuran Calls. *Symmetry*, 11(3), 40. <https://doi.org/10.3390/sym11030405>

**Luque, A., Romero-Lemos, J., Carrasco, A., y Barbancho, J.** (2018a). Improving Classification Algorithms by Considering Score Series in Wireless Acoustic Sensor Networks. *Sensors*, 18(8), 2465. <https://doi.org/10.3390/s18082465>

**Luque, A., Romero-Lemos, J., Carrasco, A., y Barbancho, J.** (2018b). Non-sequential automatic classification of anuran sounds for the estimation of climate-change indicators. *Expert Systems with Applications*, 95(1), 248-260. <https://doi.org/10.1016/j.eswa.2017.11.016>