

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 667 626**

21 Número de solicitud: 201600960

51 Int. Cl.:

**G10L 15/00** (2013.01)

**G10L 15/02** (2006.01)

**G10L 15/10** (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

11.11.2016

43 Fecha de publicación de la solicitud:

11.05.2018

Fecha de concesión:

21.03.2019

45 Fecha de publicación de la concesión:

28.03.2019

73 Titular/es:

**UNIVERSIDAD DE SEVILLA (100.0%)  
Secretariado de Transferencia del conocimiento y  
emprendimiento Paseo de las Delicias s/n -  
pabellón de Brasil  
41013 Sevilla (Sevilla) ES**

72 Inventor/es:

**CARRASCO MUÑOZ , Alejandro;  
LUQUE SANDRA, Amalia;  
ROMERO LEMOS , Javier y  
BARBANCHO CONCEJERO, Julio**

54 Título: **Sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas**

57 Resumen:

La presente invención tiene por objeto un sistema de identificación de sonidos que se basa en la descripción y selección de unos pocos parámetros caracterizadores de los mismos, la obtención de series derivadas de puntuaciones de clasificación de los mismos, y la asignación definitiva a una clase de sonido mediante la caracterización paramétrica y la clasificación de la serie derivada. La invención se encuadra en el sector de la tecnología electrónica y de las comunicaciones, concretamente de aplicación en sistemas de proceso de información y sistemas de archivo y recuperación, entre otros.

ES 2 667 626 B1

Aviso: Se puede realizar consulta prevista por el art. 37.3.8 LP 11/1986.

## DESCRIPCIÓN

Sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas

### Objeto de la invención

- 5 La presente invención tiene por objeto un sistema de identificación de sonidos que se basa en la descripción y selección de unos pocos parámetros caracterizadores de los mismos, la obtención de series derivadas de puntuaciones de clasificación de los mismos, y la asignación definitiva a una clase de sonido mediante la caracterización paramétrica y la clasificación de la serie derivada. La invención se encuadra en el
- 10 sector de la tecnología electrónica y de las comunicaciones, concretamente de aplicación en sistemas de proceso de información y sistemas de archivo y recuperación, entre otros.

### Estado de la técnica

- 15 El primer paso en la identificación de sonidos consiste en la extracción de las características del mismo, es decir, en la obtención un conjunto de parámetros que lo representen. Estos parámetros normalmente toman la forma de un vector que evoluciona en el tiempo. Para su obtención se pueden utilizar procesos temporales, espectrales, homomórficos, de codificación predictiva lineal, etc. Un resumen de las
- 20 técnicas de extracción de características de sonidos puede encontrarse en [1],[2],[3].

- Muchos de los procedimientos de extracción de características de un sonido dividen el mismo en fragmentos temporales (ventanas) de una duración muy reducida (típicamente unas pocas centésimas de segundo). De cada una de esas ventanas se obtienen unos pocos parámetros que las caracterizan, estando muy extendido el uso
- 25 de los definidos en la norma MPEG-7 [4] o los derivados de los coeficientes de Mel: MFCCs [5]. Es típico que el número de parámetros extraídos esté en torno a 20.

- Frecuentemente, a la fase de extracción le sigue un proceso de construcción de características, es decir, de obtención de parámetros derivados que reflejan comportamientos adicionales del sonido (o de la ventana de sonido). Un conjunto
- 30 típico de parámetros contruidos son aquellos que tratan de representar la evolución temporal del sonido, entre los que destacan, por ejemplo, las diferencias de primer y

segundo orden [6],[7]; o el uso conjunto de parámetros de ventanas consecutivas mediante la técnica de las ventanas deslizantes [8].

Los procesos anteriores pueden proporcionar vectores de características con un número elevado de parámetros (varios centenares), lo que incrementa sensiblemente los tiempos de proceso necesarios para la clasificación de sonidos. Además, la relevancia de cada parámetro en la tarea de clasificación puede ser muy dispar. Es pues habitual que el paso siguiente en el proceso de identificación de sonidos sea la selección de características, es decir, la obtención de un subconjunto de parámetros lo más reducido posible sin que afecte sensiblemente a la capacidad de clasificación posterior. Un resumen de las técnicas utilizadas en la selección de características puede encontrarse en [10],[11]. Dentro de éstas, las técnicas de filtro [12],[13] suelen ser las que presentan una mejor eficiencia computacional.

Una vez extraídas, construidas y seleccionadas las características, se utilizan los vectores resultantes para identificar los sonidos. Para ello se emplean distintas técnicas de clasificación que comparan las características de los sonidos con las de uno o varios patrones. En [14],[15] pueden encontrarse un buen resumen de las técnicas más habitualmente utilizadas. Entre ellas destacan las basadas en modelos ocultos de Markov [16] que es, además, la técnica recomendada en la norma MPEG-7.

La invención plantea un sistema novedoso y eficiente de identificar un sonido, es decir, de reconocerlo como perteneciente a una clase dentro de un conjunto predeterminado de ellas.

La solución propuesta parte de la extracción de características de ventanas de sonido, basadas en parámetros normalizados MPEG-7 seguida de la construcción de parámetros mediante ventanas deslizantes.

Las características generadas se utilizan para clasificar cada ventana con alguna técnica estándar de minería de datos (árboles de decisión, clasificador bayesiano, ...).

Tanto la utilización de parámetros MPEG-7 como su clasificación por minería de datos han sido ya descritos en la literatura técnica y no forman parte las reivindicaciones de la patente.

30

#### Referencias

[1] Lu, L., & Hanjalić, A. (2009). *Audio Representation*. In *Encyclopedia of Database Systems* (pp. 160-167). Springer US.

- [2] Sharan, R. V., & Moir, T. J. (2016). *An Overview of Applications and Advancements in Automatic Sound Recognition*. *Neurocomputing*.
- [3] Cowling, M., & Sitte, R. (2003). *Comparison of techniques for environmental sound recognition*. *Pattern recognition letters*, 24(15), 2895-2907.
- 5 [4] ISO (2001). *ISO/IEC FDIS 15938-4:2001: Information Technology — Multimedia Content Description Interface —Part 4: Audio*
- [5] Young, S., Evermann, G., & Gales, M. (2012). *The HTK book (ZOL)*.
- [6] Sharma, S., Shukla, A., & Mishra, P. (2014). *Speech and Language Recognition using MFCC and DELTA-MFCC*. *International Journal of Engineering Trends and*  
 10 *Technology (IJETT)*, 12(9), 449-452
- [7] Hossan, M. A., Memon, S., & Gregory, M. A. (2010, December). *A novel approach for MFCC feature extraction*. In *Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on* (pp. 1-5). IEEE
- [8] Chu, C. S. J. (1995). *Time series segmentation: A sliding window approach*. *Information Sciences*, 85(1), 147-173  
 15
- [9] Beniwal, S., & Arora, J. (2012). *Classification and feature selection techniques in data mining*. *International Journal of Engineering Research & Technology (IJERT)*, 1(6).
- [10] Guyon, I., & Elisseeff, A. (2003). *An introduction to variable and feature selection*.  
 20 *The Journal of Machine Learning Research*, 3, 1157-1182.
- [11] Liu, H., & Motoda, H. (1998). *Feature extraction, construction and selection: A data mining perspective*. Springer Science & Business Media
- [12] Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (Eds.). (2008). *Feature extraction: foundations and applications* (Vol. 207). Springer.
- 25 [13] Liu, H., Hussain, F., Tan, C. L., & Dash, M. (2002). *Discretization: An enabling technique*. *Data mining and knowledge discovery*, 6(4), 393-423.
- [14] Aggarwal, C. C. (2007). *Data streams: models and algorithms* (Vol. 31). Springer Science & Business Media.
- [15] Fu, T. C. (2011). *A review on time series data mining*. *Engineering Applications of*  
 30 *Artificial Intelligence*, 24(1), 164-181.
- [16] Rabiner, L. R. (1989). *A tutorial on hidden Markov models and selected applications in speech recognition*. *Proceedings of the IEEE*, 77(2), 257-286.

### 35 Descripción de las figuras

Figura 1.- Diagrama del proceso de clasificación de sonidos objeto de la invención.

**Descripción de la invención**

La presente invención tiene por objeto un sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas que comprende las siguientes etapas:

- 5 a. Obtención de series derivadas,  $p_{ik}$ , a partir de un clasificador de sonido que asigna una puntuación  $p_{ik}$  que mide la cercanía de cada ventana  $k$  a cada clase de  $i$ -ésimo sonido.
- b. Caracterización de cada serie derivada  $p_{ik}$ , siendo considerada cada serie derivada, como ventana única de sonido de la que se obtienen un  
10 conjunto de parámetros MPEG-7.
- c. Selección de características reduciendo el número de parámetros MPEG-7 que representa a cada serie derivada a partir de la distancia de Jensen-Shanon corregida.
- d. Identificación del sonido aplicando técnicas estándar de minería de  
15 datos a los parámetros MPEG-7 seleccionados.

La novedad de la invención consiste en los dos siguientes aspectos:

- Selección de características mediante Jensen-Shanon
  - Las varias decenas de características generadas son reducidas por un método de selección basado en la distancia de Jensen-Shanon  
20 corregida
- Clasificación del sonido a partir de series derivadas
  - El clasificador de ventanas asigna una puntuación (normalmente una probabilidad) para cada ventana y cada clase de sonido. Esto produce un conjunto de series temporales "derivadas", tantas como clases de  
25 sonido se deseen clasificar
  - Cada una de las series derivadas es caracterizada también mediante parámetros MPEG-7, considerando el conjunto de la serie como una única ventana
  - La clasificación del sonido se realiza aplicando alguna técnica estándar  
30 de minería de datos a los parámetros MPEG-7 de la serie derivada

Una vez obtenidas y seleccionadas las características de cada ventana de un sonido, el método de clasificación propuesto en la invención es el siguiente:

1. A cada ventana  $k$ -ésima, se le aplica una técnica de clasificación que asigna una puntuación  $p_{ik}$  que mide la cercanía (normalmente la probabilidad) de esa ventana a cada clase  $i$ -ésima.
2. Analizando la totalidad de las ventanas del sonido, para cada clase  $i$ -ésima se obtiene una serie temporal de puntuaciones  $p_i$ , o serie derivada.
3. Cada clase  $i$ -ésima de puntuaciones  $p_i$  se trata como si fuese una ventana única de sonido (no se segmenta) y de ella se obtienen un conjunto de parámetros MPEG-7
4. Se reduce el número de parámetros que representa a cada serie derivada mediante el mismo método de selección de características basado en la distancia de Jensen-Shanon corregida que se detalla más adelante.
5. Al conjunto del sonido, representado ahora por un pequeño conjunto de parámetros, se le aplica una técnica de clasificación que permite la identificación definitiva con alguna de las clases predeterminadas.

#### *Método de selección basado en la distancia de Jensen-Shanon*

El primer paso del método de selección de características propuesto en la invención consiste en determinar la separabilidad de clases de sonido atendiendo a cada uno de los parámetros. Para el cálculo del índice de separabilidad de clases  $\Psi_\theta$  atendiendo al parámetro  $\theta$ -ésimo, se procede de la siguiente forma:

1. Para cada clase  $i$ -ésima, se obtienen los valores del parámetro  $\theta$ -ésimo en todas las ventanas de sonido pertenecientes a dicha clase.
2. Para cada clase  $i$ -ésima, se calculan las funciones de densidad de probabilidad  $f_{\theta_i}$  de los valores del parámetro  $\theta$ -ésimo en las ventanas de sonido pertenecientes a dicha clase.
3. Para cada posible par de clases  $i$  y  $j$  se calcula la divergencia de Jensen-Shanon que viene dada por

$$D_{JS}(f_{\theta_i}, f_{\theta_j}) = \frac{1}{2} \int_{-\infty}^{\infty} f_{\theta_i} \log_2 \frac{2f_{\theta_i}}{f_{\theta_i} + f_{\theta_j}} dx + \frac{1}{2} \int_{-\infty}^{\infty} f_{\theta_j} \log_2 \frac{2f_{\theta_j}}{f_{\theta_i} + f_{\theta_j}} dx$$

4. Para cada posible par de clases  $i$  y  $j$  se calcula la distancia de Jensen-Shanon que viene dada por

$$d_{JS}(f_{\theta_i}, f_{\theta_j}) \equiv \sqrt{D_{JS}(f_{\theta_i}, f_{\theta_j})}$$

5. Se calcula el índice de separabilidad  $\Psi_{\theta}$  como

$$\Psi_{\theta} \equiv \sqrt{\prod_{i=1}^{n-1} \prod_{j=i+1}^n d_{JS}(f_{\theta_i}, f_{\theta_j})}$$

expresión en la que  $n$  es el número de clases de sonidos que se pretenden identificar; y  $N$  es el número de distancias de Jensen-Shanon calculadas cuyo valor viene dado por

$$N = \frac{n(n-1)}{2}$$

5

*Método de selección basado en la distancia de Jensen-Shanon corregida*

El método de selección de características propuesto en la invención utiliza el índice de separabilidad calculado en el apartado anterior, pero lo corrige en función de la correlación entre parámetros. El proceso propuesto es el siguiente:

- 10 1. Se calcula la matriz de correlación parámetro-parámetro  $e_p$ , que está formada por los elementos  $e_{ij}$  que representan la correlación entre los parámetros  $i$  y  $j$ . Estos valores se calculan como

$$e_{ij} = \frac{\text{cov}(x_i, x_j)}{\sqrt{\text{var}(x_i)\text{var}(x_j)}} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}}$$

expresión en la que  $x_{ik}$  representa el valor del parámetro  $i$ -ésimo en la ventana  $k$ -ésima;  $\bar{x}_i$  es el valor medio del parámetro  $i$ -ésimo; y  $m$  es el número total de  
15 ventanas.

2. A partir de la matriz de correlación  $e_p$  se calcula la matriz de independencia entre parámetros definida como

$$\Delta \equiv 1 - |e_p|$$

3. Para cada parámetro se calcula el índice de separabilidad  $\Psi_i$  siguiendo el procedimiento apuntado en el apartado anterior.

4. Se elige como parámetro más relevante aquél con el valor mayor de  $\Psi_i$ . Se incorpora al conjunto (ordenado) de parámetros relevantes  $\mathcal{R}$  y se elimina del conjunto de parámetros pendientes de analizar  $\mathcal{P}$ .
5. Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la independencia  $\delta_{jk}$  con respecto a cada uno de los  $k$ -ésimos parámetros en  $\mathcal{R}$ .
6. Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la independencia mínima con respecto a  $\mathcal{R}$  definida como

$$\mu_j \equiv \min_k \delta_{jk}$$

7. Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la separabilidad corregida, definida como

$$\kappa_j \equiv \Psi_j \cdot \mu_j$$

8. Se elige como parámetro más relevante aquél con el valor mayor de  $\kappa_j$ . Se incorpora al conjunto (ordenado) de parámetros relevantes  $\mathcal{R}$  y se elimina del conjunto de parámetros pendientes de analizar  $\mathcal{P}$ .
9. Se iteran los pasos 5 al 8 hasta haber analizado todos los parámetros:  $\mathcal{P} = \emptyset$

- 15 Con respecto al estado de la técnica, la invención propuesta plantea tres ventajas principales:

- La identificación de sonidos mediante clasificación de series derivadas mejora sensiblemente la tasa de éxito en la clasificación de sonidos
- La reducción del número de parámetros en la caracterización de las ventanas de sonido hace sensiblemente más rápido el proceso de clasificación
- El método de selección de características propuesto permite una más rápida determinación de un subconjunto de parámetros que representan al sonido para su posterior clasificación.

25

#### **Modo de realización de la invención**

Una realización posible del sistema propuesto en la invención es la siguiente:

1. Se registra un conjunto de sonidos, por ejemplo, utilizando una frecuencia de muestreo de 44.1 kHz y 16 bits de resolución

2. Se eligen un subconjunto de sonido (por ejemplo un 10%) que se consideren significativos de cada una de las clases que se desean identificar. Este subconjunto constituirán los sonidos patrón.
3. En cada patrón se determinan los segmentos que son identificativos de la clase que representan.
4. Cada sonido es segmentado en ventanas de corta duración, por ejemplo, de 10 milisegundos.
5. De cada ventana se obtienen diferentes parámetros MPEG-7. El conjunto de ellos puede ser, por ejemplo, el siguiente:
  - Potencia total
  - Potencia relevante, esto es, la comprendida en una determinada banda de frecuencias, por ejemplo, entre 500 Hz. y 5kHz.
  - Centroide de potencia
  - Dispersión espectral
  - Planitud
  - Tono
  - Razón de armonicidad
  - Frecuencia límite de armonicidad
  - Frecuencia de los 3 primeros formantes
  - Ancho de banda de los 3 primeros formantes
  - Centroide armónico
  - Desviación armónica
  - Dispersión armónica
  - Variación armónica
6. Mediante el método de selección de características propuesto en la invención, se reduce el número de parámetros eligiendo, por ejemplo, los 5 más significativos.
7. Para cada ventana, se construyen otras características mediante la técnica de ventana deslizante. Es decir, la ventana se identifica con los 5 parámetros propios más los 5 correspondientes a las ventanas colindantes. El tamaño de la ventana deslizante puede ser, por ejemplo de 5. Se caracteriza pues cada ventana mediante 25 (5x5) parámetros.
8. Cada ventana es clasificada por comparación entre sus características y las características de las ventanas de los patrones. El método de clasificación puede ser, por ejemplo, el del árbol de decisión. Este clasificador genera una

puntuación de la cercanía de cada ventana a cada clase (probabilidad de pertenecer a esa clase).

9. La aplicación del clasificador a la secuencia de las ventanas de un sonido produce una serie derivada (puntuaciones) para cada una de las clases de sonido.

5

10. Cada una de las series derivadas es considerada como si fuese una ventana única de un sonido y de ella se extraen los parámetros MPEG-7 expresados anteriormente.

11. Mediante el sistema de selección de características propuesto en la invención, se reduce el número de parámetros con los que se caracteriza cada serie derivada eligiendo, por ejemplo, los 5 más significativos. Si tenemos, por ejemplo, 10 clases, las series derivadas de cada sonido se caracterizan mediante 50 (5x10) parámetros.

10

12. Cada sonido es clasificado por comparación entre las características de sus series derivadas y las características de las series derivadas de los patrones. El método de clasificación puede ser, por ejemplo, el del árbol de decisión.

15

20

25

**Reivindicaciones**

1. Sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas caracterizado porque comprende:

- 5 e. Obtención de series derivadas,  $p_i$ , a partir de un clasificador de sonido que asigna una puntuación  $p_k$  que mide la cercanía de cada ventana  $k$  a cada clase de  $i$ -ésimo sonido.
- f. Caracterización de cada serie derivada  $p_i$ , siendo considerada cada serie derivada, como ventana única de sonido de la que se obtienen un conjunto de parámetros MPEG-7.
- 10 g. Selección de características reduciendo el número de parámetros MPEG-7 que representa a cada serie derivada a partir de la distancia de Jensen-Shanon corregida.
- h. Identificación del sonido aplicando técnicas estándar de minería de datos a los parámetros MPEG-7 seleccionados.

2. Sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas caracterizado porque la selección de características de una ventana de sonido a partir de la distancia de Jensen-Shanon se obtiene calculando el índice de separabilidad de clases  $\psi_c$  atendiendo al parámetro MPEG-7  $\theta$ -ésimo que comprende:

- i) Para cada clase  $i$ -ésima, se obtienen los valores del parámetro  $\theta$ -ésimo en todas las ventanas de sonido pertenecientes a dicha clase.
- ii) Para cada clase  $i$ -ésima, se calculan las funciones de densidad de probabilidad  $f_{\theta_i}$  de los valores del parámetro  $\theta$ -ésimo en las ventanas de sonido pertenecientes a dicha clase.
- 25 iii) Para cada posible par de clases  $i$  y  $j$  se calcula la divergencia de Jensen-Shanon  $D_{JS}$  que viene dada por

$$D_{JS}(f_{\theta_i}, f_{\theta_j}) = \frac{1}{2} \int_{-\infty}^{\infty} f_{\theta_i} \log_2 \frac{2f_{\theta_i}}{f_{\theta_i} + f_{\theta_j}} dx + \frac{1}{2} \int_{-\infty}^{\infty} f_{\theta_j} \log_2 \frac{2f_{\theta_j}}{f_{\theta_i} + f_{\theta_j}} dx$$

- iv) Para cada posible par de clases  $i$  y  $j$  se calcula la distancia de Jensen-Shanon  $d_{JS}$  que viene dada por

$$d_{JS}(f_{\theta_i}, f_{\theta_j}) \equiv \sqrt{D_{JS}(f_{\theta_i}, f_{\theta_j})}$$

v) El cálculo del índice de separabilidad  $\Psi_\theta$  como

$$\Psi_\theta \equiv \sqrt{N \prod_{i=1}^{n-1} \prod_{j=i+1}^n d_{JS}(f_{\theta_i}, f_{\theta_j})}$$

expresión en la que  $n$  es el número de clases de sonidos que se pretenden identificar, y  $N$  es el número de distancias de Jensen-Shanon calculadas cuyo valor viene dado por

$$N = \frac{n(n-1)}{2}$$

5

3. Sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas según reivindicaciones anteriores caracterizado porque la selección de características de una ventana de sonido a partir de la distancia de Jensen-Shanon descrita en la reivindicación 2 es corregida en función de la correlación entre parámetros MPEG-7 y que comprende:

10

i) El cálculo de la matriz de correlación parámetro-parámetro  $e_p$ , que está formada por los elementos  $e_{ij}$  que representan la correlación entre los parámetros  $i$  y  $j$ . Estos valores se calculan como

15

$$e_{ij} = \frac{\text{cov}(x_i, x_j)}{\sqrt{\text{var}(x_i)\text{var}(x_j)}} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}}$$

expresión en la que  $x_{ik}$  representa el valor del parámetro  $i$ -ésimo en la ventana  $k$ -ésima;  $\bar{x}_i$  es el valor medio del parámetro  $i$ -ésimo; y  $m$  es el número total de ventanas.

ii) A partir de la matriz de correlación  $e_p$  se calcula la matriz de independencia entre parámetros definida como

20

$$\Delta \equiv 1 - |e_p|$$

iii) Para cada parámetro se calcula el índice de separabilidad  $\Psi_i$  siguiendo el procedimiento apuntado en el apartado anterior.

iv) Se elige como parámetro más relevante aquél con el valor mayor de  $\Psi_i$ . Se incorpora al conjunto (ordenado) de parámetros relevantes  $\mathcal{R}$  y se elimina del conjunto de parámetros pendientes de analizar  $\mathcal{P}$ .

25

v) Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la independencia  $\delta_{jk}$  con respecto a cada uno de los  $k$ -ésimos parámetros en  $\mathcal{R}$ .

vi) Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la independencia mínima con respecto a  $\mathcal{R}$  definida como

$$\mu_j \equiv \min_k \delta_{jk}$$

5 vii) Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la separabilidad corregida, definida como

$$\kappa_j \equiv \Psi_j \cdot \mu_j$$

viii) Se elige como parámetro más relevante aquél con el valor mayor de  $\kappa_j$ . Se incorpora al conjunto (ordenado) de parámetros relevantes  $\mathcal{R}$  y se elimina del conjunto de parámetros pendientes de analizar  $\mathcal{P}$ .

10 ix) Se iteran los pasos v) a viii) hasta haber analizado todos los parámetros:  $\mathcal{P} = \emptyset$

4. Sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas caracterizado porque la selección de características de una serie derivada a partir de la distancia de Jensen-Shanon se obtiene calculando el índice de separabilidad de clases  $\psi_\theta$  atendiendo al parámetro MPEG-7  $\theta$ -ésimo que comprende:

i) Para cada clase  $i$ -ésima, se obtienen los valores del parámetro  $\theta$ -ésimo en todas las ventanas de sonido pertenecientes a dicha clase.

15 ii) Para cada clase  $i$ -ésima, se calculan las funciones de densidad de probabilidad  $f_{\theta_i}$  de los valores del parámetro  $\theta$ -ésimo en las ventanas de sonido pertenecientes a dicha clase.

20 iii) Para cada posible par de clases  $i$  y  $j$  se calcula la divergencia de Jensen-Shanon  $D_{JS}$  que viene dada por

$$D_{JS}(f_{\theta_i}, f_{\theta_j}) = \frac{1}{2} \int_{-\infty}^{\infty} f_{\theta_i} \log_2 \frac{2f_{\theta_i}}{f_{\theta_i} + f_{\theta_j}} dx + \frac{1}{2} \int_{-\infty}^{\infty} f_{\theta_j} \log_2 \frac{2f_{\theta_j}}{f_{\theta_i} + f_{\theta_j}} dx$$

25 iv) Para cada posible par de clases  $i$  y  $j$  se calcula la distancia de Jensen-Shanon  $d_{JS}$  que viene dada por

$$d_{JS}(f_{\theta_i}, f_{\theta_j}) \equiv \sqrt{D_{JS}(f_{\theta_i}, f_{\theta_j})}$$

v) Se calcula el índice de separabilidad  $\Psi_\theta$  como

$$\Psi_{\theta} \equiv \sqrt{\prod_{l=1}^{n-1} \prod_{j=l+1}^n d_{JS}(f_{\theta_l}, f_{\theta_j})}$$

expresión en la que  $n$  es el número de clases de sonidos que se pretenden identificar; y  $N$  es el número de distancias de Jensen-Shanon calculadas cuyo valor viene dado por

$$N = \frac{n(n-1)}{2}$$

5

5. Sistema de identificación de sonidos mediante clasificación paramétrica de series derivadas según reivindicaciones 1 y 4 caracterizado porque la selección de características de una serie derivada a partir de la distancia de Jensen – Shanon descrita en la reivindicación 4 es corregida en función de la correlación entre parámetros MPEG-7 y que comprende

10

i) El cálculo de la matriz de correlación parámetro-parámetro  $e_p$ , que está formada por los elementos  $e_{ij}$  que representan la correlación entre los parámetros  $i$  y  $j$ . Estos valores se calculan como

$$e_{ij} = \frac{cov(x_i, x_j)}{\sqrt{var(x_i)var(x_j)}} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}}$$

15

expresión en la que  $x_{ik}$  representa el valor del parámetro  $i$ -ésimo en la ventana  $k$ -ésima;  $\bar{x}_i$  es el valor medio del parámetro  $i$ -ésimo; y  $m$  es el número total de ventanas.

20

ii) A partir de la matriz de correlación  $e_p$  se calcula la matriz de independencia entre parámetros definida como

$$\Delta \equiv 1 - |e_p|$$

iii) Para cada parámetro se calcula el índice de separabilidad  $\Psi_i$  siguiendo el procedimiento apuntado en el apartado anterior.

iv) Se elige como parámetro más relevante aquél con el valor mayor de  $\Psi_i$ . Se incorpora al conjunto (ordenado) de parámetros relevantes  $\mathcal{R}$  y se elimina del conjunto de parámetros pendientes de analizar  $\mathcal{P}$ .

25

v) Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la independencia  $\delta_{jk}$  con respecto a cada uno de los  $k$ -ésimos parámetros en  $\mathcal{R}$ .

5 vi) Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la independencia mínima con respecto a  $\mathcal{R}$  definida como

$$\mu_j \equiv \min_k \delta_{jk}$$

vii) Para cada uno de los  $j$ -ésimos parámetros en  $\mathcal{P}$ , se calcula la separabilidad corregida, definida como

$$\kappa_j \equiv \Psi_j \cdot \mu_j$$

10 viii) Se elige como parámetro más relevante aquél con el valor mayor de  $\kappa_j$ . Se incorpora al conjunto (ordenado) de parámetros relevantes  $\mathcal{R}$  y se elimina del conjunto de parámetros pendientes de analizar  $\mathcal{P}$ .

ix) Se iteran los pasos v) a viii) hasta haber analizado todos los parámetros:  $\mathcal{P} = \emptyset$

15

20

25

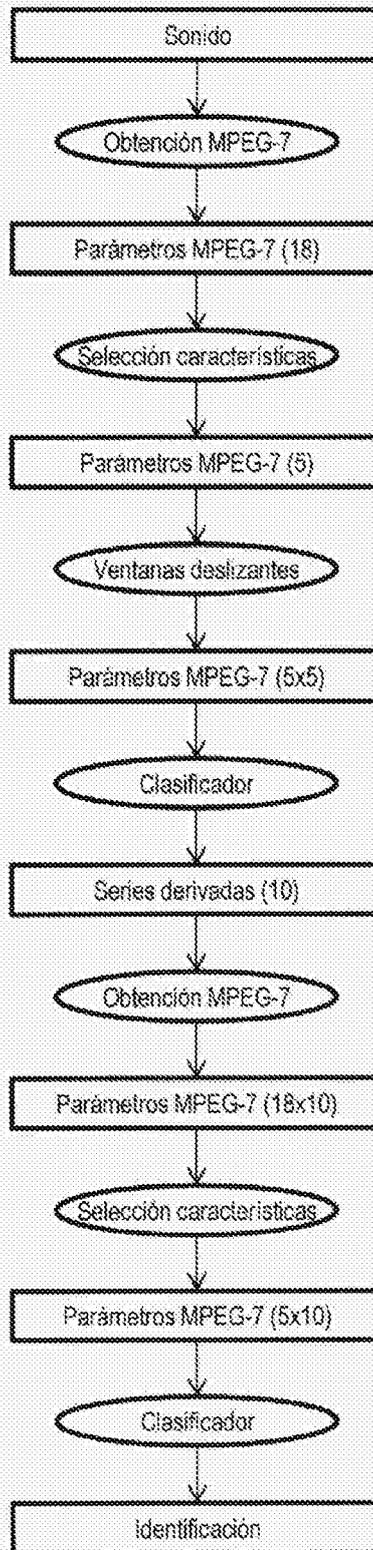


Figura 1



- ②<sup>1</sup> N.º solicitud: 201600960  
②<sup>2</sup> Fecha de presentación de la solicitud: 11.11.2016  
③<sup>2</sup> Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤ Int. Cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	⑥ Documentos citados	Reivindicaciones afectadas
Y	Mark Levy et al.. STRUCTURAL SEGMENTATION OF MUSICAL AUDIO BY CONSTRAINED CLUSTERING. 01/02/2008, Vol. 16, Páginas 318 - 326, ISSN 1558-7916, <DOI: 10.1109/TASL.2007.910781>. Todo el documento	1, 2, 4
A		3, 5
Y	Xiangsheng Huang et al. JENSEN-SHANNON BOOSTING LEARNING FOR OBJECT RECOGNITION. 20/06/2005, Vol. 2, Páginas 144 - 149, ISBN 978-0-7695-2372-9 ; ISBN 0-7695-2372-2, <DOI:10.1109/CVPR.2005.197>. Todo el documento	1, 2, 4
A		3, 5
A	Barranco-López V et al. ENTROPIC TEXTURE-EDGE DETECTION FOR IMAGE SEGMENTATION. 25/05/1995, Vol. 31, Páginas 867 - 869, ISSN 0013-5194, <DOI: 10.1049/el:19950598>. Todo el documento	1-5
A	He M et al.. BAND SELECTION BASED ON FEATURE WEIGHTING FOR CLASSIFICATION OF HYPERSPECTRAL DATA. 01/04/2005, Vol. 2, Páginas 156 - 159, ISSN 1545-598X, <DOI: 10.1109/LGRS.2005.844658>. Todo el documento	1-5
A	Adrian Silvescu et al.. COMBINING SUPER-STRUCTURING AND ABSTRACTION ON SEQUENCE CLASSIFICATION. 06/12/2009, Páginas 986 - 991, ISBN 978-1-4244-5242-2 ; ISBN 1-4244-5242-2. Todo el documento	1-5

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº:

Fecha de realización del informe  
07.09.2017

Examinador  
M. Muñoz Sanchez

Página  
1/6

## CLASIFICACIÓN OBJETO DE LA SOLICITUD

**G10L15/00** (2013.01)

**G10L15/02** (2006.01)

**G10L15/10** (2006.01)

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G10L

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, NPL, XPI3E, XPIEE

Fecha de Realización de la Opinión Escrita: 07.09.2017

#### Declaración

<b>Novedad (Art. 6.1 LP 11/1986)</b>	Reivindicaciones 1-5	<b>SI</b>
	Reivindicaciones	<b>NO</b>
<b>Actividad inventiva (Art. 8.1 LP11/1986)</b>	Reivindicaciones 3,5	<b>SI</b>
	Reivindicaciones 1,2,4	<b>NO</b>

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

#### Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

**1. Documentos considerados.-**

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	Mark Levy et al.. STRUCTURAL SEGMENTATION OF MUSICAL AUDIO BY CONSTRAINED CLUSTERING. IEEE Transactions on Audio, Speech and Language Processing, 20080201 IEEE. Vol. 16, Nº 2, Páginas 318 - 326, ISSN 1558-7916, <DOI:10.1109/TASL.2007.910781>	01.02.2008
D02	Xiangsheng Huang et al.. JENSEN-SHANNON BOOSTING LEARNING FOR OBJECT RECOGNITION. Proceedings / 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005 : [20 - 25 June 2005, San Diego, CA], 20050620; 20050620 - 20050626 IEEE, Piscataway, NJ, USA. Vol. 2, Páginas 144 - 149, ISSN ISBN 978-0-7695-2372-9; ISBN 0-7695-2372-2, <DOI:10.1109/CVPR.2005.197>	20.06.2005
D03	Barranco-López V et al. ENTROPIC TEXTURE-EDGE DETECTION FOR IMAGE SEGMENTATION. ELECTRONICS LET, 19950525 IEE STEVENAGE, GB. Vol. 31, Nº 11, Páginas 867 - 869, ISSN 0013-5194, <DOI:10.1049/el:19950598>	25.05.1995
D04	He M et al.. BAND SELECTION BASED ON FEATURE WEIGHTING FOR CLASSIFICATION OF HYPERSPECTRAL DATA. IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, 20050401 IEEE SERVICE CENTER, NEW YORK, NY, US. Vol. 2, Nº 2, Páginas 156 - 159, ISSN 1545-598X, <DOI:10.1109/LGRS.2005.844658>	01.04.2005
D05	Adrian Silvescu et al.. COMBINING SUPER-STRUCTURING AND ABSTRACTION ON SEQUENCE CLASSIFICATION. Ninth IEEE International Conference on Data Mining, 2009 (ICDM &apos;09), 20091206 IEEE, Piscataway, NJ, USA. Páginas 986 - 991, ISSN ISBN 978-1-4244-5242-2; ISBN 1-4244-5242-2	06.12.2009

**2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración**

Se considera D01 el documento más próximo del estado de la técnica al objeto de la solicitud.

**Reivindicaciones independientes**

**Reivindicación 1:** El documento D01 divulga un método de segmentación de secuencias de sonidos musicales a partir de características espectrales (parámetros en la solicitud) en un entorno MPEG-7 en el que se analiza un conjunto de características. Dicho conjunto de características luego se reduce utilizando el método de las componentes principales. También se hace mención a la divergencia de Jensen-Shannon a la hora de evaluar similitud entre histogramas (ventana en la solicitud). Combinando histogramas de instantáneas (*frames*) vecinas se agrupan en *clusters* que definen los segmentos. Las diferencias entre el documento D01 y esta reivindicación se refieren a:

- La utilización de una distancia Jensen-Shannon corregida
- La identificación en sí del sonido aplicando técnicas de minería de datos

El efecto técnico de las diferencias consistiría en que la selección de parámetros sería más eficiente y así la identificación más rápida.

El documento D02 por su parte presenta un sistema de reconocimiento de objetos basado en la selección de parámetros utilizando la distancia (divergencia) de Jensen-Shannon, eligiéndose iterativamente como parámetros más discriminantes y eficientes aquellos que maximizan dicha distancia.

En dicho documento, no se menciona el uso concreto de una variante de la distancia Jensen-Shannon, característica esta que en el sentido genérico con el que se expresa en la reivindicación 1 es una alternativa evidente para el experto en la materia como se ilustra por ejemplo en el documento D03 (*Some variants of the above procedure may also be implemented, using a sliding window W divided into four equal adjacent subwindows. The overall divergence JS4, is computed from the above general definition, as well as different partial divergences by taking the above quadrant distributions in pairs. Several combinations result, from which the most appropriate for a particular problem may be chosen*), por ejemplo, la combinación de distancias entre pares de distribuciones.

En consecuencia, la combinación de los documentos D01 y D02 afecta a la actividad inventiva de la reivindicación 1 según el art. 8.1 de la Ley 11/86 de Patentes.

### Reivindicaciones dependientes

Reivindicaciones 2, 4: Como se ha indicado anteriormente la combinación de pares de distribuciones es una posibilidad conocida para el experto en la materia y, en particular, las operaciones sencillas para combinarlas como la suma o el producto resultarían evidentes para él.

Por tanto, la combinación de los documentos D01 y D02 afecta a la actividad inventiva de las reivindicaciones 2, 4 según el art. 8.1 de la Ley 11/86 de Patentes.

Reivindicaciones 3, 5: la utilización la correlación entre parámetros solo se menciona brevemente en el documento D04 (*only choose bands with correlation coefficients below a pre-specified threshold T*) y en este no se menciona siquiera el uso de la distancia de Jensen-Shannon. Por tanto, ninguno de los documentos anteriores divulga la corrección reivindicada de la separabilidad ni existe información en ellos que, sin aplicar actividad inventiva, orientara al experto en la materia a, partiendo de ellos, obtener la solución al problema técnico de cómo mejorar todavía más la selección de parámetros y conseguir así una mayor rapidez (mayor reducción de la dimensión del espacio a explorar) en el análisis de patrones para la identificación de sonidos.

Por tanto, las reivindicaciones 3, 5 tienen actividad inventiva según el art. 8.1 de la Ley 11/86 de Patentes.