

Estudio de sensibilidad de las características para la detección de eventos acústicos utilizando máquinas de vectores soporte

PACS: 43.60.Np Técnicas acústicas de procesamiento de señales para las redes neuronales y sistemas de aprendizaje

Martín Morató,Irene; Cobos,Máximo; Ferri,Francesc J; Jaume Segura.
Universitat de Valencia, Avda Universitat s/n, Burjassot, Valencia, (España), Tel: 963543211, E-Mail: etse@uv.es

Palabras Clave: Clasificación de eventos acústicos, máquinas de vectores soporte, selección de características, audio semántico.

ABSTRACT.

Automatic recognition of multiple acoustic events is an interesting problem in machine listening that generalizes the classical speech/non-speech or speech/music classification problem. Classification is usually performed by means of hidden Markov models (HMMs) or support vector machines (SVMs) considering traditional sets of features based on Mel-frequency cepstral coefficients (MFCCs) and their temporal derivatives. However, while these features are routinely used by many systems, it is not yet understood which their relative importance in the classification task is. This paper presents a preliminary study to assess the sensitivity of these features under a common SVM framework, aiming at providing deeper insight into appropriate low-level audio event representation for classification tasks.

RESUMEN.

El reconocimiento acústico de múltiples eventos es un problema interesante en audición por computación que generaliza clásico problema de clasificación de habla/no habla o habla/música. Típicamente la clasificación se realiza utilizando modelos ocultos de Markov (HMM) o máquinas de vectores soporte (SVM) considerando conjuntos tradicionales de características basados en coeficientes cepstrales de Mel (MFCCs) y sus derivadas temporales. Sin embargo, mientras estas características se emplean de forma rutinaria por muchos sistemas, todavía no se entiende cuál es su importancia en la tarea de clasificación. Este trabajo presenta un estudio preliminar de la sensibilidad de estas características bajo un marco común de SVM, con el objetivo de proporcionar una visión a más bajo nivel de la representación de los eventos acústicos para la tarea de clasificación.

1. INTRODUCCIÓN

La clasificación de eventos acústicos en ambientes sonoros reales ha atraído recientemente la atención de las comunidades de procesamiento de señal y reconocimiento de patrones por sus múltiples aplicaciones en las ciudades inteligentes (Smart cities), monitorización ambiental o vida cotidiana asistida por el entorno [1,2]. Como en muchos otros problemas de reconocimiento de patrones, el rendimiento general de los sistemas de reconocimiento de eventos acústicos se basa en la fase de extracción de características, la fase de clasificación y su apropiada combinación.

En éste contexto, los modelos ocultos de Markov (HMMs) [3,4] y las máquinas de vectores soporte [5,6] han sido ampliamente utilizados para este tipo de tareas como sistemas de clasificación válidos. Motivados desde los sistemas de reconocimientos de voz automático, los conjuntos de características ampliamente utilizados incluyen los coeficientes cepstrales en las frecuencias de Mel (MFCCs) y la energía según bancos de filtros en la escala de Mel, calculadas a partir de ventanas cortas de tiempo sobre la señal [7]. Normalmente se incluye información temporal concatenando sus derivadas temporales de primer orden (normalmente conocida como características Δ).

Sin embargo, mientras que esas características han demostrado ser útiles en varios sistemas, pocos trabajos han investigado su importancia dentro del rendimiento general del sistema [8]. Por otra parte, el mismo conjunto de características pueden ser calculadas considerando diferentes parámetros (ancho de banda, número de bandas de frecuencia, componentes seleccionados, etc...) [9], lo que no ayuda a entender el impacto de esas características de un trabajo a otro.

Este trabajo presenta un estudio preliminar enfocado a estudiar la influencia de estas características ampliamente utilizadas para calcular el rendimiento de clasificación utilizando un marco común de SVM. Con este fin, se ha diseñado un sistema de referencia para la clasificación de eventos acústicos siguiendo el enfoque uno contra todos, entrenando 11 clasificadores binarios para el reconocimiento de 11 clases de eventos acústicos grabados en un entorno real de oficina. El impacto de cada característica en el rendimiento general de la clasificación será estudiado eliminando una a una su influencia del conjunto de datos de entrenamiento, observando la sensibilidad del sistema para cada una de las características individualmente de forma independiente [10,11]. Utilizando este procedimiento, se crea una ordenación según el impacto de las diferentes características en el rendimiento general del sistema, ofreciendo así una visión interesante de la dependencia del sistema para cada una de las características individuales.

2. SISTEMA DE REFERENCIA

El sistema de referencia considerado en este trabajo consiste en una etapa de extracción de características y en una etapa de clasificación. Mientras algunos planteamientos incluyen otra etapa de pre-procesado para aumentar la solidez del sistema, como la utilización de algoritmos de reducción de ruido, nosotros preferimos evitar el uso de algoritmos específicos para concentrarnos en el comportamiento del clasificador y la extracción de características. De forma similar, gran parte de los sistemas de reconocimiento de eventos están diseñados para realizar la segmentación de la señal acústica de entrada en eventos sonoros individuales antes de la clasificación. Dado que este se puede considerar un problema separado del de la clasificación, nosotros asumimos también que los eventos sonoros están perfectamente aislados del flujo de entrada, de ese modo, evitamos errores que puedan venir de la fase de segmentación. Las siguientes subsecciones describen específicamente el sistema de referencia en términos de características, diseño del clasificador y los conjuntos de datos de entrenamiento y test.

2.1. Características

Las características son extraídas mediante un análisis de la señal a corto plazo utilizando ventanas Hamming de 25 ms de duración y 10 ms de solapamiento, con una frecuencia de muestreo de 44.1 kHz. Para cada ventana temporal, se extraen las siguientes características (entre paréntesis su número):

1. MFCCs (20): Extraemos 21 MFCCs, descartando el primer coeficiente DCT. El ancho de banda considerado en el análisis va de 64 Hz a 8 kHz, con un banco de filtros según la escala de Mel compuesto por 31 filtros triangulares. Un filtro de pre-acentuación con coeficiente 0.97 se aplica a la señal previa a su análisis de frecuencia.
2. Energías de bancos de filtros según la escala de Mel (MFEs) (31): Las energías de los anteriores bancos de filtros se guardan como características adicionales (antes de aplicar la habitual decorrelación DCT a los MFCCs).
3. Δ MFCCs (20): En general, los coeficientes Δ se calculan como:

$$d_t = \frac{\sum_{n=1}^N n(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^N n^2},$$

Donde d_t es un coeficiente delta de la ventana t calculada, en términos de los coeficientes estáticos c_{t+n} a c_{t-n} , con $N = 2$. En el caso de $d_t = \Delta MFCCs$, los coeficientes c_t son cualquiera de los MFCCs calculados.

4. $\Delta MFEs$ (31): Como en el caso anterior, cuando $d_t = \Delta MFEs$, los diferentes coeficientes c_t se corresponden con los MFEs calculados.

Adicionalmente, dado que distintos eventos tienen diferente duración, se divide la longitud total del evento en tres secciones temporales uniformes, correspondiendo a una parte de inicio ($T=1$), una parte intermedia ($T=2$) y una parte final ($T=3$), de estas secciones se calcula la media y la desviación estándar para cada una de las características arriba mencionadas. Como resultado, obtenemos un vector agregado de características con dimensiones $(20 \times 2 \times 31 \times 2) \times 3 \times 2 = 612$ para cada uno de los ejemplos.

2.2. Clasificador

El clasificador multi-clase consiste en 11 clasificadores SVM binarios entrenados siguiendo el método uno contra todos, donde cada clasificador binario es entrenado considerando ejemplos de la clase de interés como ejemplos positivos y los ejemplos del resto de clases como ejemplos negativos. Cada clasificador binario aplica un kernel de funciones de base radial (RBF) eligiendo el parámetro gamma y su coste mediante un proceso de búsqueda intensivo. Para clasificar un evento individual, las características correspondientes a un evento en concreto se introducen en los 11 clasificadores binarios previamente entrenados, obteniendo 11 probabilidades a posteriori correspondientes a la clase (positiva) de interés. La clase ganadora se selecciona como la correspondiente a la probabilidad a posteriori más alta. Las características extraídas se normalizan para tener media cero y varianza unitaria antes de introducirlas en el clasificador. A pesar de su simplicidad, esta clasificación ofrece resultados muy razonables en comparación con otros enfoques del estado del arte y se ha demostrado que es apropiada para nuestro objetivo de analizar las características.

2.3. Base de Datos

La base de datos de entrenamiento utilizada en este trabajo es la proporcionada por el DCASE 2016 IEEE (Tarea 2, detección de eventos sonoros en audio sintético), la cual contiene eventos de entrenamiento de 11 clases: alarma, aclarar garganta, toser, golpear la puerta, cerrar una puerta, cajón, risa, teclado, llaves (dejar sobre la mesa), girar una página, teléfono y habla. La base de datos contiene un total de 220 instancias (20 por clase) y es la utilizada para entrenar todos los experimentos. En cuanto a los datos de test disponemos de tres pistas de audio de similar dificultad, de los cuales extraemos los eventos individuales perfectamente aislados utilizando las anotaciones que nos proporciona el DCASE. Cada una de las pistas de test se compone de un total de 33 eventos acústicos (3 por clase), cada pista será referenciada como *test1*, *test2* y *test3*.

3. ORDENACIÓN DE CARACTERÍSTICAS

Obtener un buen ordenamiento o ranking de los datos es un problema muy interesante en el aprendizaje automático, como se puede seleccionar una apropiada representación de los datos de forma reducida para una tarea determinada. Los algoritmos de selección clásicos [12] se pueden utilizar para este propósito. Sin embargo, incluso el conjunto de algoritmos secuenciales más rápidos y simples necesitan un número cuadrático de modelos de entrenamiento y evaluaciones para obtener el ranking de características correspondiente.

Una alternativa interesante y conveniente es la llamada poda basada en sensibilidad (SBP, Sensitivity Based Prunning) [10]. Esta fue propuesta dentro del contexto de los modelos de redes neuronales donde el entrenamiento de un número cuadrático de modelos puede ser prohibitivo. Para obtener un ranking de características conveniente, el SBP calcula un único modelo de entrenamiento y un número lineal de evaluaciones parciales para calcular el promedio de sensibilidad para cada característica. La medida de sensibilidad se calcula de la siguiente forma.

Sea $g(\mathbf{x})$ la función/clasificador y $\mathbf{x} = [x_1, \dots, x_d]^T$ su correspondiente vector de características de entrada. Sea M_g una medida particular de rendimiento en g utilizando un conjunto de entrenamiento dado. La sensibilidad para la i -ésima característica es dada por la siguiente fórmula

$$S_M(i) = M_g - M_{g_i}$$

Donde $g_i(\mathbf{x}) = g([x_1, \dots, \bar{x}_i, \dots, x_d]^T)$ y \bar{x}_i es el promedio de la i -ésima entrada en el conjunto de entrenamiento.

En nuestra particular implementación, los predictores SVM aprenderán del conjunto de entrenamiento utilizando un esquema de validación cruzada utilizando 3 conjuntos (3-fold CV) y se tendrán en cuenta dos medidas de rendimiento: el CV estimado de la precisión en clasificación y la media de la probabilidad a posteriori asignada a cada muestra por el SVM correspondiente a su verdadera clase. El primer criterio, el cual utiliza el número de aciertos, S_A , corresponde exactamente al que se usará después para calcular el rendimiento final en los datos de test, pero ofrece una baja resolución dado que hay muy pocos datos de entrenamiento disponibles. Por otra parte, el criterio que utiliza probabilidades, S_p no guarda mucha relación con el rendimiento final (alta probabilidad a posteriori positiva no garantiza necesariamente la más alta entre todas las clases SVM) pero permite una evaluación más sutil cuando se utiliza como criterio de selección de características.

3.1. Ranking de Características por Bloques

Una pequeña modificación del procedimiento de ranking de características anteriormente descrito se obtiene agrupando las características en bloques significativos y calculando la sensibilidad para cada grupo como un todo. Los grupos engloban características de la misma familia así como mismas secciones temporales, resultando en un total de 24 bloques diferentes:

- 6 MFE, y 6 Δ MFE bloques, con 31 características cada uno.
- 6 MFE, y 6 Δ MFE bloques, con 20 características cada uno.

Cada uno de los 6 bloques de arriba, corresponden a 3 valores medios y su correspondiente desviación estándar para $T = 1, 2, 3$ respectivamente.

4. EXPERIMENTOS

La experimentación llevada a cabo en este trabajo se ha realizado utilizando algoritmos estándar para entrenar todos los modelos SVM. La ordenación de características se ha obtenido utilizando solo datos de entrenamiento. En particular, se ha utilizado un esquema de 3-fold CV para obtener medidas de sensibilidad promediadas y todo el proceso se ha repetido 4 veces utilizando diferentes particiones aleatorias. La calibración de los parámetros se ha realizado para cada uno de los conjuntos de forma automatizada como se ha explicado en la sección 2.2, por último las medidas hacen referencia al promedio de las cuatro repeticiones.

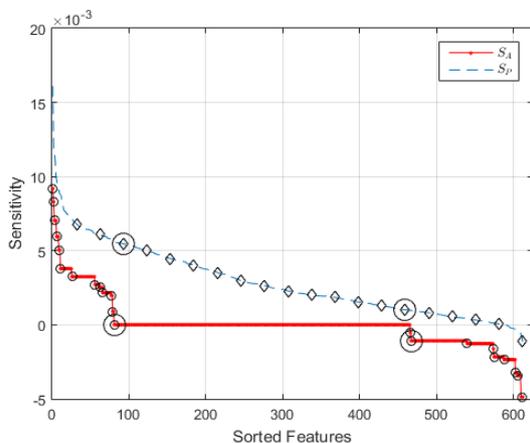


Fig. 1: Valores de sensibilidad en orden descendente para cada característica individual.

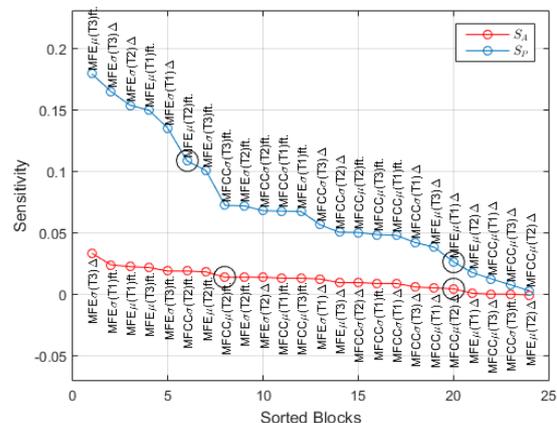


Fig. 2: Valores de sensibilidad por bloques en orden descendente.

4.1. Análisis de Sensibilidad

Los valores resultantes de sensibilidad utilizando precisión de clasificación, S_A , y el promedio de las probabilidades a posteriori, S_P , se muestran en la Fig. 1. Como se puede observar, S_A , toma solamente algunos valores diferentes, mientras que S_P muestra un comportamiento más suave. Por otra parte, ambos criterios conducen a una región central con poca variación comparada a la variación presente en los extremos. Para poder realizar una evaluación más detallada, se seleccionan un subconjunto de 22 y 20 características de las curvas S_A y S_P , respectivamente. En el primer caso éstas se corresponden aproximadamente a los tamaños más pequeños para un mismo valor de sensibilidad y se muestran en la Fig. 1 con círculos pequeños. En el segundo caso, los 20 tamaños (representados como diamantes pequeños) se han obtenido de forma uniforme a lo largo de todo el rango de posibles valores. Dos tamaños específicos para cada curva se han marcado con grandes círculos para futura discusión en la sección 4.3.

De forma similar, Fig.2 muestra la sensibilidad cuando consideramos bloques de características utilizando también los dos criterios S_A y S_P . Las curvas muestran 24 valores diferentes, ordenados de mayor a menor sensibilidad. En este caso, se muestran las etiquetas para indicar el orden específico de cada bloque. Los resultados en bloques para S_P muestran mayor sensibilidad que para el caso S_A , especialmente para los bloques ordenados primero (del 1 al 7). Curiosamente, ambos criterios tienden a mostrar que el rendimiento es más sensible para las características MFE que para las características MFCC. En cualquier caso, más adelante se mostrarán una serie de experimentos para elaborar más sobre este aspecto.

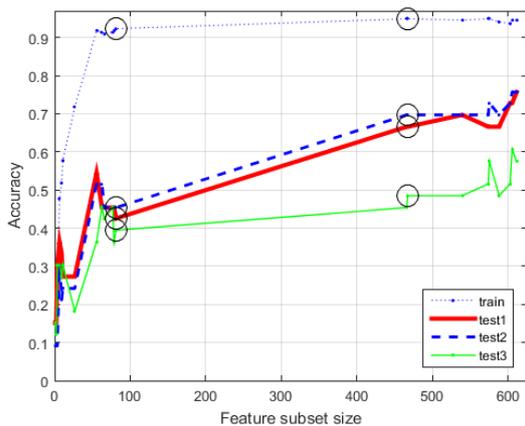


Fig. 3: Rendimiento obtenido usando la secuencia del subconjunto cuando las características están ordenadas según S_A . Las curvas muestran estimaciones CV de precisión usando los datos de entrenamiento (train) y test (test1, test2, test3).

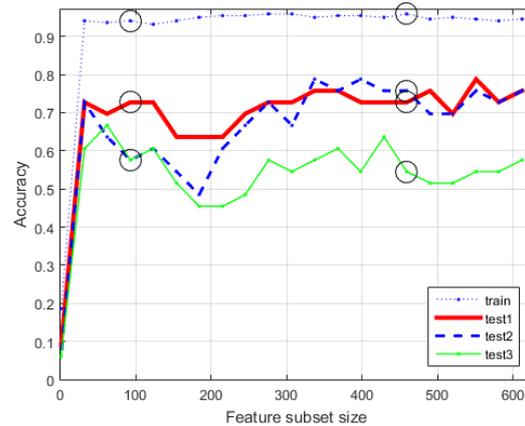


Fig. 4: Rendimiento obtenido usando la secuencia del subconjunto cuando las características están ordenadas según S_P . Las curvas muestran estimaciones CV de precisión usando los datos de entrenamiento (train) y test (test1, test2, test3).

4.2. Precisión vs. Dimensionalidad

Después del análisis de sensibilidad, el objetivo es ahora mostrar como el rendimiento de clasificación se degrada a la vez que se reduce la dimensión a lo largo de las correspondientes secuencias de subconjuntos identificadas para cada criterio. Para ello, se han diseñado dos experimentos diferentes de aprendizaje y clasificación. Por un lado, 3-fold CV estimaciones de precisión se han obtenido utilizando datos de entrenamiento. Por otro lado, los modelos SVM finales se han construido utilizando todos los datos de entrenamiento y se han utilizado para clasificar las tres pistas de audio para test disponibles. Todos estos resultados de clasificación se muestran en las Figs. 3 y 4 para cada secuencia de subconjuntos, respectivamente.

Como se esperaba, las medidas de precisión obtenidas con CV muestran un comportamiento suave y consistente al incremento de dimensionalidad en ambas figuras cuando se considera el conjunto de datos de entrenamiento. En ambos casos, para este conjunto de datos, la precisión tiende a aumentar hasta alcanzar un tamaño de subconjunto de 81 características ordenadas. Desde este tamaño hasta un subconjunto de tamaño 467, el rendimiento crece poco a poco para ambos criterios. Sin embargo, aunque se pueden observar tendencias similares para el conjunto de datos de prueba, las curvas de precisión muestran una estructura más variable, especialmente en el caso de S_p . Hay que tener en cuenta que los resultados para *test1* y *test2* son similares en ambos casos, pero *test3* muestra una precisión menor. Para S_p , los conjuntos de datos *test2* y *test3* también muestran una profunda disminución alrededor de un tamaño de subconjunto de 93 que se recupera después de 460.

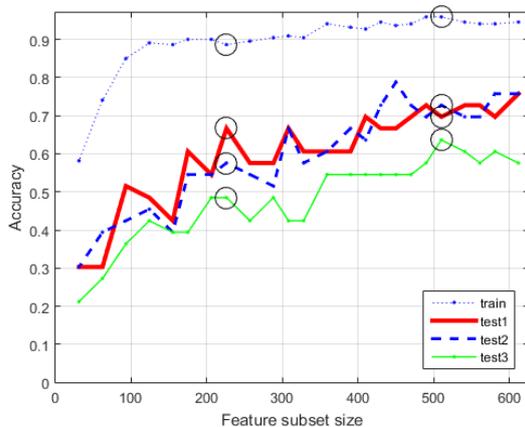


Fig. 5: Rendimiento obtenido cuando se usa la secuencia del subconjunto por bloques ordenados según S_A .

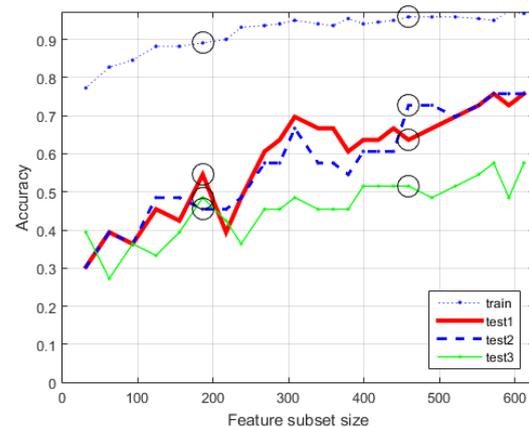


Fig. 6: Rendimiento obtenido cuando se usa la secuencia del subconjunto por bloques ordenados según S_p .

Los resultados del mismo experimento pero utilizando bloques de características se muestran en las Fig. 5 y 6. Como en el caso de las características individuales el rendimiento tiende a incrementar, pero en una proporción menor. Esto se debe al hecho de que los subconjuntos de características se han obtenido agregando características provenientes de la misma familia, donde algunas de ellas pueden no ser interesantes para nada desde el punto de vista de la sensibilidad.

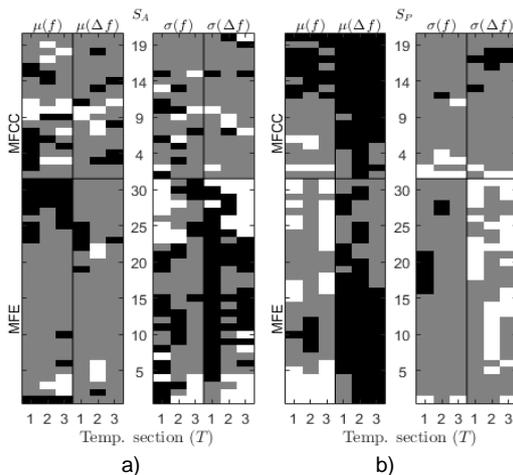


Fig. 7: Subconjuntos de características utilizando a) S_A con tamaños 81 y 467, y b) S_p con tamaños 93 y 459. Las características en blanco/negro son las más/menos sensibles y las grises corresponden a valores intermedios de sensibilidad.

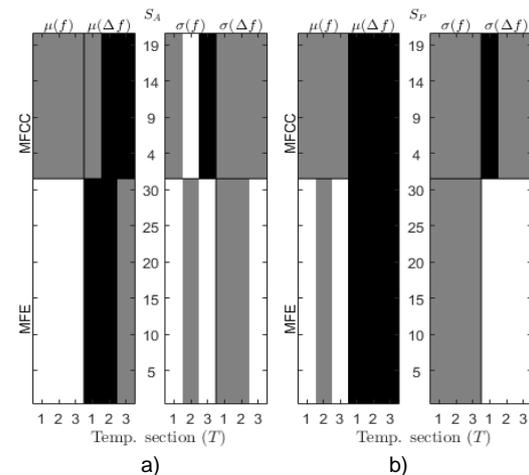


Fig. 8: Subconjuntos de características utilizando a) S_A con tamaños 226 y 510, y b) S_p con tamaños 186 y 460. Las características en blanco/negro son las más/menos sensibles y las grises corresponden a valores intermedios de sensibilidad.

Además, los resultados del conjunto de entrenamiento son más suaves que los provenientes del conjunto de test, donde, otra vez, se obtienen curvas similares para *test1* y para *test2* y los peores valores de precisión para *test3*.

4.3. Mapas de Sensibilidad de Características

Dos tamaños representativos de subconjuntos de características se han elegido y se han marcado con grandes círculos en las diferentes curvas en la Fig. 1 a 6. Para ver gráficamente cuales son las características más importantes de acuerdo a cada criterio, se han realizado mapas de características mostrando tres niveles diferentes de importancia. En particular, en la Fig. 7a las 81 mejores características según el criterio S_A se muestran en blanco. Las siguientes mejores hasta sumar 467 se muestran en gris y las restantes hasta 612 se muestran en negro. Exactamente lo mismo se muestra en la Fig. 7b para tamaños 93 y 460 ordenados según S_p . Cada mapa consiste en dos submapas (valores de media, $\mu(\cdot)$ y de desviación estándar, $\sigma(\cdot)$, respectivamente) organizados como sigue. Características (f) y características Δ (Δf) se muestran a la parte izquierda/derecha de cada submapa mientras que MFCCs y MFE se muestran a la parte arriba/bajo ambos separados por unas finas líneas negras. Las etiquetas 1, 2 y 3 bajo de cada mapa indican a que columna hace referencia a la primera, segunda o última sección temporal del evento.

Mapas similares para los bloques de características se muestran en Fig. 8. Los subconjuntos elegidos para los bloques de análisis son [226, 510] para la sensibilidad S_A y [186, 460] para la sensibilidad S_p . Hay que tener en cuenta que los subconjuntos no son iguales a los utilizados para las características individuales porque los límites tienen que coincidir con el inicio y final de los bloques. Sin embargo, estas pequeñas diferencias no afectan a la comparación de los mapas obtenidos en los dos métodos (Sensibilidad de las características individuales y sensibilidad de las características por bloques).

5. CONCLUSIONES

De los mapas en las Fig. 7 y 8, queda claro que de forma global las características Δ son menos importantes que sus correspondientes características. Sin embargo, es posible distinguir algunas muy importantes a lo largo de $\sigma(\Delta f)$. Otro hecho global algo inesperado es que de forma general las características MFE son más importantes que las MFCCs. Moviéndonos a la comparación entre características seleccionadas por cada criterio, observamos que las mejores 93 características en el caso de S_p (Fig. 7b) están mejor concentradas en MFE en concreto para el caso de $\mu(f)$ y $\sigma(\Delta f)$. Al mismo tiempo, las peores características se concentran en la parte de las medias de MFCC. Para el caso de S_A (Fig. 7a), las características más importantes están dispersas alrededor de todos los submapas pero siguen existiendo características más importantes en las regiones de MFE. De hecho, cuando se observa el mapa del bloque de sensibilidad en la Fig. 8 se observa claramente que los bloques de MFE son claramente más importantes que los bloques MFCC. Solo para las medias de las características Δ MFE la sensibilidad tiende a ser más baja que para los correspondientes bloques MFCC. Curiosamente, este hecho también parece sostenerse cuando se estudian los mapas de la sensibilidad de características individuales.

Es interesante analizar la sensibilidad obtenida por secciones temporales. Las características correspondientes a la primera y última parte del evento ($T = 1,3$) tienden a ser más importantes considerandos ambos mapas de sensibilidad individual y por bloques. Solo en pocos casos la sección intermedia ($T=2$) parece tener importancia. Esto se puede deber a que son eventos muy cortos e impulsivos, por lo que los cambios se concentran en el inicio y final del evento.

En resumen, a la vez que existen diferencias entre criterios de sensibilidad, parece que los MFE de forma global tienen un impacto más grande en el rendimiento de clasificación que los

MFCCs. De forma similar, la importancia de las características Δ tiende a ser menor que su correspondientes valor de características.

En términos de estadística, parece estar bastante balanceada la relevancia de las medias y las desviaciones estándar, aunque las calculadas a partir de la sección temporal intermedia tienen significativamente un impacto menor en el rendimiento de clasificación. Por último, los autores desean remarcar que se debe tener especial cuidado en la generalización de los resultados de este trabajo, el cual, solo ha considerado un conjunto específico (y limitado, en términos de número de ejemplos) como base de datos de audio para un entorno acústico dado. En futuros trabajos se desea estudiar diferentes dominios de audio, como entornos acústicos urbanos, clasificación del hablante o clasificación de géneros musicales. Adicionalmente, otros esquemas de clasificación, como los basados en HMM o redes neuronales artificiales se tendrán en consideración.

6. REFERENCIAS

- [1] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1733–1746, Oct. 2015.
- [2] A. Mesaros, T. Heittola, A. Eronen, and T. Virtanen, "Acoustic event detection in real life recordings," in *Signal Processing Conference, 2010 18th European*, Aug. 2010, pp. 1267–1271.
- [3] A. J. Eronen, V. T. Peltonen, J. T. Tuomi, A. P. Klapuri, S. Fagerlund, Timo Sorsa, G. Lorho, and Jyri Huopaniemi, "Audio-based context recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 321–329, Jan. 2006.
- [4] M. E. Niessen, T. L. M. Van Kasteren, and A. Merentitis, "Hierarchical modeling using automated sub-clustering for sound event recognition," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2013)*, Oct. 2013, pp. 1–4.
- [5] A. Temko and C. Nadeu, "Acoustic event detection in meeting-room environments," *Pattern Recogn. Lett.*, vol. 30, no. 14, pp. 1281–1288, 2009.
- [6] A. Temko and J. Nadeu, C. and Biel, "Acoustic event detection: Svm-based system and evaluation setup in CLEAR'07," in *Multimodal Technologies for Perception of Humans*, pp. 354–363. Springer, 2008.
- [7] L. Rabiner and B. Juang, *Fundamentals of speech recognition*, Prentice Hall, 1993.
- [8] Xiaodan Zhuang, Xi Zhou, T. S. Huang, and M. Hasegawa-Johnson, "Feature analysis and selection for acoustic event detection," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, Mar. 2008, pp. 17–20.
- [9] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of MFCC," *J. of Comput. Sci. and Tech.*, vol. 16, no. 6, pp. 582–589, 2001.
- [10] J.E. Moody and J. Utans, "Principled architecture selection for neural networks: Application to corporate bond rating prediction," in *Advances in Neural Information Processing Systems*, J.E. Moody, S.J. Hanson, and R.P. Lippmann, Eds., vol. 4, pp. 683–690. Morgan Kauffmann Publishers, 1992.
- [11] W. Díaz-Villanueva, F. J. Ferri, and V. Cerverón, "Learning improved feature rankings through decremental input pruning for support vector based drug activity prediction," in *Trends in Applied Intelligent Systems*, pp. 653–661. Springer, 2010.
- [12] Anil Jain and Douglas Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, 1997.