TECNIACÚSTICA
CUENCA 2023

DETECCIÓN AUTOMÁTICA DE GOTEOS A PARTIR DE MODELOS SINTÉTICOS DE SU HUELLA ACÚSTICA.

Manuel A. Sobreira Seoane
AtlanTTic-Research Center for Telecommunication Technologies.
Universidad de Vigo

RESUMEN

Una de las utilizades de las técnicas de inteligencia artificial es la detección y clasificación de sonidos que pueden generar señales de alerta. En este artículo se trata con la detección de sonidos impulsivos (goteos) en entornos domésticos. La detección de sonidos generados por agua es de gran interés en entornos domésticos, al permitir detectar la posible aparición de fugas. En el caso de goteos es muy difícil obtener una base de datos que contenga la gran cantidad de casuísticas sonoras que se pueden dar en la realidad. Pensemos que el sonido generado por un goteo va a depender de multitud de factores que harán que la huella sonora del sonido generado presente grandes variaciones en función del tipo de superficie sobre la cae la gota (madera, plancha de acero, suelo cerámico, etc.), del tamaño de la gota, etc. Para generar una base de datos que recoja la gran variabilidad de la vida real, se ha partido de un conjunto de goteos reales, que se ha expandido mediante la utilización de modelos sintéticos de goteos. Se generaron tablas de entrenamiento utilizando características como la variación temporal de los coeficientes Mel (DMFCC), la kurtisis y la densidad de probabilidad de las altas frecuencias, que se emplearon para alimentar un clasificador SVM. Para la evaluación de las prestaciones del clasificador se utilizó un conjunto de prueba de goteos reales sobre superficies no contempladas en la generación.

ABSTRACT

In this paper, the problem of detection of an specific event– a drip–in the domestic acoustic scene is addressed. The detection of noises generated by water can be of great interest in domestic scenes because it can help to prevent domestic floods. In the case of drip sounds, it is quite difficult to get real sounds covering the different sound qualities that different drops may have. Their sound depends on many factors as their size, the kind of surface they hit (wood, steel), etc. One of the problems when using Artificial Intelligence techniques to detect and classify sounds is the lack of data to train the models. In order to approach real life as much as possible, a database including real sounds has been expended using synthesized. A training set has been set up using the speed of variation of the MFCC, the kurtosis and the probability density function of the high frequencies as features to train a SVM classifier. The performance of the classifier has been tested using a data set containing recordings of real drops.

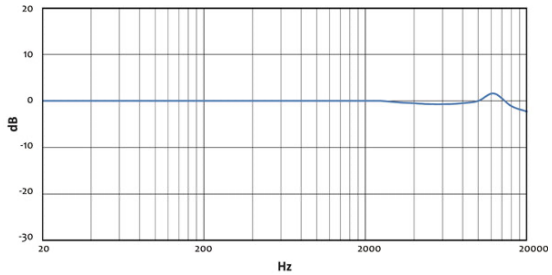
Palabras Clave— detección de sonidos, SVM, clasificación de sonidos, inteligencia artificial, características acústicas.

1. INTRODUCCIÓN

En los últimos años la detección automática de eventos acústicos está mostrando un gran incremento en el interés y actividad. Son cada vez más numerosas las publicaciones donde se tratan con la detección de sonidos en diversos entornos. Por ejemplo, Alsina et al [1] aplica los MFCC – *Mel Frequency Cepstral Coefficients* – y un clasificador SVM – *Support Vector Machine* – a la detección de múltiples sonidos en el entorno doméstico con el objetivo de monitorizar a personas de edad avanzada. En Sharma et al [2,3] se analizan distintas características acústicas, como la variación del tono o la envolvente espectral (formantes), para detectar la razón por la que llora un bebé. Jian [4] procesa el espectrograma normalizado para obtener características como la entropía. La clasificación de eventos acústicos en aplicaciones reales es una tarea desafiante y no exenta de dificultades, debido a menudo a la dificultad para obtener bases de datos con suficientes ítems y a la influencia de ruidos de fondo variables. Definimos un evento acústico, como un sonido de duración limitada en el tiempo, que por sus características se distingue del entorno acústico: el llanto de un bebé, un goteo, una puerta que se abate, etc. Los eventos acústicos en entornos reales aparecerán mezclados con diversos sonidos que formarán parte de la escena acústica.

Un sonido doméstico o una escena acústica pueden contener sonidos como música, habla, un zumbido de lavavajillas, un ruido de aspiradora, etc. En [5], se describen las estrategias y métodos a seguir para diseñar un sistema capaz de detectar y clasificar eventos en escenarios complejos. En este trabajo, se aborda el problema de la detección de un evento específico, un goteo, en la escena acústica doméstica. Para diseñar un sistema simple, se evita la necesidad de una posible separación de la fuente, con el fin de extraer el evento de interés del ruido de fondo, centrándose en la extracción de características capaces de detectar el evento específico inmerso en ruido de fondo.

(a) Micrófono i436



(b) Respuesta en frecuencia

Figura 1. Micrófono omnidireccional i436. Sensibilidad S= 44 dB (6.3 mV/Pa) y S/N> 63 dB (fuente, especificaciones del fabricante en http://www.mic-w.com/).

La detección de ruidos generados por el agua puede ser de gran interés en escenas domésticas porque puede ayudar a prevenir inundaciones: la detección de un simple goteo puede ayudar a detectar un problema en su etapa inicial, antes de tener repercusiones mayores.

En la sección 2 de este artículo, se describe el método que se ha utilizado para generar la base de datos de entrenamiento. En la sección 3 se especifican las características acústicas utilizadas para entrenar un clasificador SVM lineal. En la sección 4, se presentan los resultados de las pruebas realizada para presentar finalmente las conclusiones del trabajo en la sección 5.

2. LA GENERACIÓN DE UNA BASE DE DATOS DE SONIDOS DE GOTEO.

Para entrenar adecuadamente un detector/clasificador de eventos acústicos se necesita una base de datos balanceada, con un conjunto suficientemente representativo de cada clase. En el caso que nos ocupa, se necesita un número suficiente de "eventos" (goteos) y un conjunto de "no eventos" que represente la variabilidad de ruidos de fondo en entornos domésticos. Para obtener la base de datos se han manejado tres fuentes de sonidos:

1. Sonidos de goteos grabados, golpeando en diversas superficies (suelo cerámico duro, madera, lavabos cerámicos, bañeras, fregaderos en aluminio y acero inoxidable). Para introducir variabilidad en la base de datos, se grabó con distintos equipos:
   - Grabaciones directas con el micrófono de un teléfono móvil.
   - Grabación desde teléfono móvil con un micrófono externo, de tipo electrect, omnidireccional, de clase 2: Mic W i 436 (ver figura 1).
   - Un Edirol estéreo R-09, con micrófono de condensador.
2. Sonidos obtenidos de la base de datos de la BBC [6] (goteos de agua y sonidos domésticos variados como aspiradoras, teteras, etc.).
3. Sonidos sintéticos: se ha recurrido a la generación de modelos sintéticos de los sonidos de goteos y mezclados con diversos ruidos de fondo domésticos.

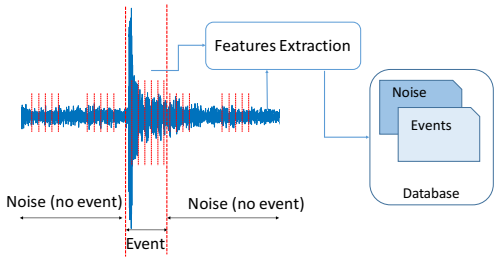


Figura 2. Extracción de características trama a trama.

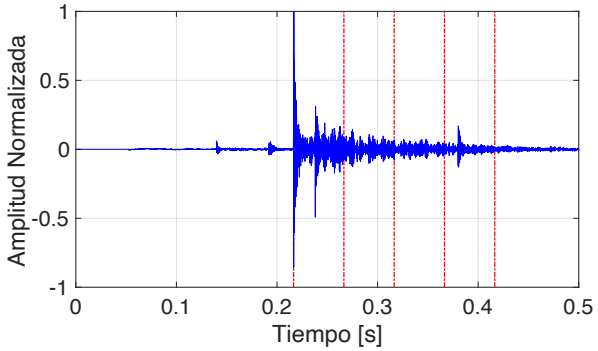


Figura 3. Señal grabada de una gota y tramas de 0.05 s.

Se ha elegido trabajar con una extracción trama a trama de las características sonoras: tal como se muestra en la figura 2, se divide la señal en tramas y se extraen sus características. Cada trama se etiqueta como "ruido de fondo" o "evento". Como el objetivo del trabajo es detectar sonidos transitorios de corta duración, el tamaño elegido de las tramas es $t_f = 50$ ms. La frecuencia de muestreo de la base de datos es $f_s = 48000$ Hz.

Tal como se muestra en la figura 3, la duración del sonido de una gota es aproximadamente de 0.2 s, lo que supone que cada evento contiene 4 tramas de la longitud

elegida. Claramente, para relaciones Señal a Ruido bajas, el sonido de una gota puede estar enmascarado por la presencia de ruido de fondo.

2.1. Expansión de la base de datos utilizando modelos sintéticos.

Tal como se ha descrito anteriormente, el propósito de este trabajo es detectar el sonido de una gota cuando esta cae sobre una superficie, bien sobre una superficie y dura y seca o sobre agua (superficie encharcada, o cavidad con agua). El sonido percibido (timbre) del sonido radiado dependerá de las características de la gota (tamaño y velocidad de caída) y de la superficie (tamaño de la superficie y características del material). Para poder entrenar el modelo de clasificación con una muestra suficientemente representativa de goteos, es necesario expandir la base de datos cambiando las características sonoras en función de la naturaleza de la superficie.

5.1.1. El sonido emitido por gotas golpeando una superficie seca.

En este caso, el sonido emitido se deberá a la radiación de la superficie golpeada por la gota. El impacto de corta duración excita las frecuencias de resonancia de la superficie, y se puede asumir que la señal emitida estará constituida por un conjunto de caídas acústicas. Se ha optado por el modelo por utilizado por Vörlander [7],

$$p(t) = \sum_{i=1}^{N}\{A_i \sin(2\pi f_i\} + \phi_i) \cdot w(t) \cdot e^{-\alpha t}, \qquad (1)$$

en el que las A_i son las amplitudes de cada caída, que dependerán del punto de impacto de la gota. Las f_i son las frecuencias de resonancia de la placa. ϕ_i son las fases iniciales asociadas a cada caída, dependientes también del punto de impacto. $w(t)$ es una ventana temporal utilizada para controlar la duración de las señales, ajustada empíricamente. Finalmente el factor α, es el amortiguamiento asociado a cada resonancia. Para simular en la base de datos que tanto el tamaño, el amortiguamiento y el punto de impacto son desconocidos a priori, se generan señales donde las frecuencias de resonancia, los factores de amortiguamiento, las fases y las amplitudes son variables aleatorias con distribución de probabilidad uniforme.

5.1.2. Goteo sobre agua.

Tal como se describe en Guyot et al [8] y Moss and Henhching Yeh [9], el sonido percibido de un goteo se debe a la generación de burbujas bajo el agua, un instante posterior al momento en el que la gota impacta la superficie de agua. Moos [9] propone algunos modelos que tratan con el sonido radiado por una burbuja con diferentes grados de complejidad. Para el presente trabajo se ha optado por el modelo del sonido generado por burbujas esféricas:

$$p(t) = \epsilon r_o \sin(2\pi f(t) \cdot t) \cdot e^{-\beta_o t},\ \epsilon \in [0.01, 0.1], \qquad (2)$$

donde ϵ es un parámetro sintonizable que permite inicializar el estado de excitación de las burbujas, r_0 es el radio de las burbujas (en metros) y:

– $f(t)$ es una frecuencia dependiente del tiempo que depende de la frecuencia de resonancia fundamental de la burbuja, f_0:

$$f(t) = fo(1 + \xi\beta_o t), \qquad (3)$$

donde ξ es un parámetro que ayuda a ajustar el efecto de incremento de frecuencia del sonido de una gota. Se toma el valor $\xi = 0.1$ como valor óptimo. La frecuencia de resonancia de una gota se calcula tal como sugiere Van Den Doel [10]:

$$f_0 = \frac{3}{r_0}, \qquad (4)$$

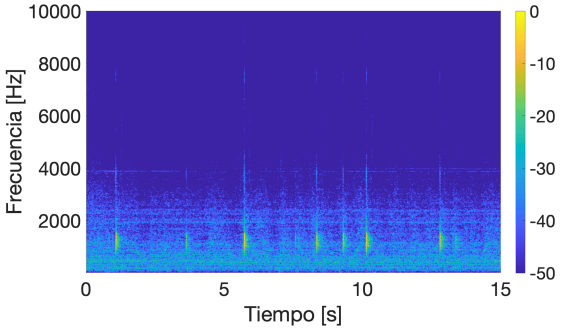
– $\beta_0 = \pi f_0 \delta_{tot}$ es el factor de atenuación de la caída exponencial, donde δ_{tot} es el amortiguamiento total de cada burbuja, que se calcula según se describe en [10]:

$$\delta_{tot} = \frac{0.13}{r_o} + 0.0072\, r_o^{-3/2}. \qquad (5)$$

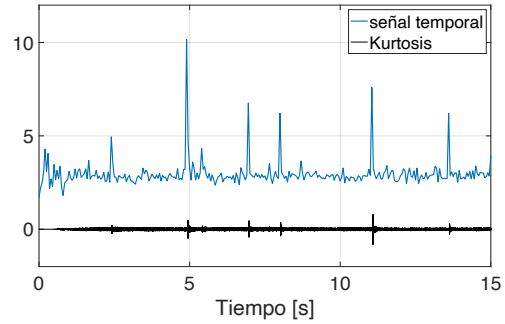
De las ecuaciones anteriores podemos deducir que el parámetro que conduce a la generación del sonido de una gota que cae sobre agua es el tamaño de esta. Tal como muestra la ecuación (2), la presión sonora generada depende del factor de atenuación y de la frecuencia de resonancia de la burbuja y ambos factores se calculan a partir del radio de la burbuja. Para expandir la base de datos con sonidos de goteos sobre agua, se ha tomado el radio como una variable aleatoria de distribución uniforme entre 0.5 y 10 mm.

2.2. La generación del conjunto de entrenamiento.

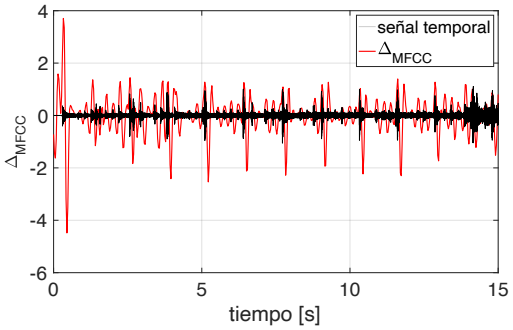
El conjunto de entrenamiento generado trata de recoger de la variabilidad de las condiciones reales. Para ello se ha generado un conjunto de 16000 tramas de goteos: 1000 tramas se corresponden con sonidos reales grabados y 15000 tramas con sonidos sintéticos generados según se ha descrito en el apartado anterior. Teniendo en cuenta que el tamaño de cada trama es de 50 ms, se ha generado una base de datos con 800 segundos de goteos. Estos sonidos se han mezclado con diversos ruidos de fondo de forma aleatoria, incluyendo distintas muestras de música de diversos géneros. De esta base de datos se extraen de forma aleatoria subconjuntos de entrenamiento, formados por unas 1000 gotas y 1000 muestras de ruido de fondo.



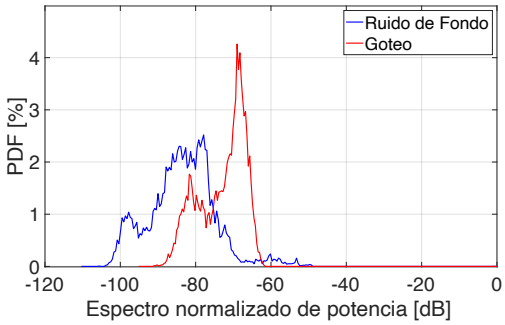
a) Espectrograma de un goteo en presencia de ruido de fondo (ruido de un aspirador doméstico).



b) kurtosis



c) Velocidad de variación de los MFCC, Δ_{MFCC}.



d) Función de densidad de probabilidad de las frecuencias altas

Figura 4. Características acústicas de 15 segundos de señal de goteo con ruido de fondo (aspiradora).

3. CARACTERÍSTICAS ACÚSTICAS

Una de las tareas críticas en la detección y clasificación es la selección de un conjunto apropiado de características: estas no deben introducir información redundante y deben maximizar la distancia entre clases [11,12]. Existe una intensiva descripción en la bibliografía de las diferentes características acústicas [13]. Las aproximaciones más comunes en la detección de eventos acústicos trasladan características que han sido utilizadas con éxito en la señal de voz a la clasificación de eventos acústicos. Por ejemplo, los coeficientes cepstrales en el escala Mel (MFCC) y otras características basadas en la representación de la señal en escala Mel (como los *logarithmic Mel-filter bank coefficients—FBANK*) son habituales en la detección de eventos acústicos [14] a pesar de la reconocida sensibilidad de estos coeficientes ante ruido de fondo [15]. Para este trabajo, teniendo en cuenta el comportamiento en tiempo y frecuencia de la señal de interés, se ha optado por las siguientes características:

- La función de densidad de probabilidad de las frecuencias altas. Tal como se observa en el espectrograma de un goteo – figura 4a, las características impulsivas de este hacen que su espectro se extienda a frecuencias superiores a los 2000 Hz, mientras que la mayoría de los ruidos domésticos concentran su energía en baja frecuencia. La figura 4 d) muestra que en el caso de un goteo es más probable obtener niveles superiores a -80 dB (espectro normalizado de potencia).
- La kurtosis: Para una trama de señal, la kurtosis se calcula como el momento estandarizado de orden 4:

$$k = \frac{E(x-\mu)^4}{\sigma^4} \qquad (6)$$

donde μ es la media y σ la desviación estándar del conjunto de muestras de la trama de señal. Si k=3, las muestras de señal estarán normalmente distribuidas, mientras que la kurtosis toma valores altos cuando dentro de la trama se presenta un transitorio. La kurtosis es muy sensible a los transitorios incluso cuando estos están enmascarados por la presencia de altos niveles de ruido de fondo. La figura 4b) muestra 15 segundos de señal y la kurtosis de un goteo en presencia de ruido de fondo. Los picos de kurtosis se corresponden con el sonido emitido por las gotas.

- Δ_{MFCC}: Se ha considerado también la velocidad de variación de los MFCC, obtenida como la diferencia

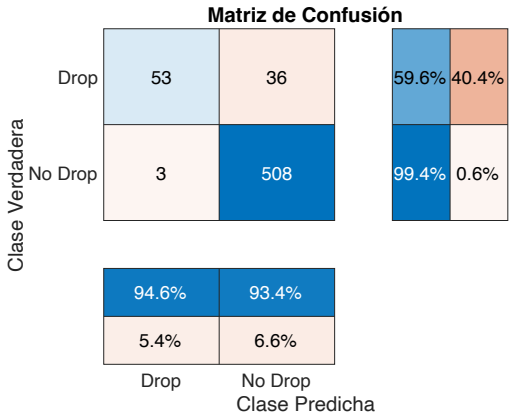
finita de orden uno entre los valores obtenidos de MFCC de dos tramas adyacentes:

$$\Delta_{MFCC}[n] = MFCC[n] - MFCC[n-1]. \quad (7)$$

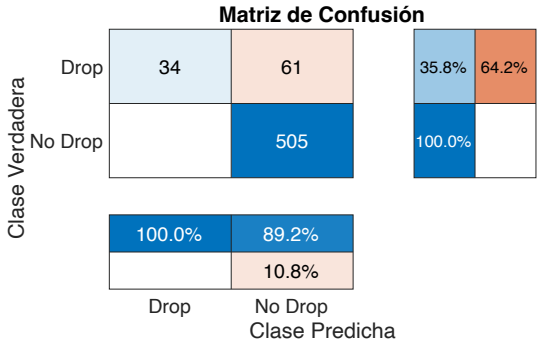
Es de esperar que en condiciones de alto ruido de fondo estacionario, la diferencia sea muy pequeña, mientras que la diferencia entre tramas será alta cuando existe un transitorio. La figura 4c) muestra la Δ_{MFCC} en el caso del ejemplo (goteo en presencia de ruido de fondo).

4. RESULTADOS Y DISCUSIÓN

Una vez seleccionado el conjunto de características a utilizar, se generan conjuntos de entrenamiento según se ha detallado anteriormente, para entrenar un clasificador lineal SVM (*Support Vector Machine)* [17,18]. Se han realizado 10 tests diferentes con diversos conjuntos de entrenamiento y test generados al azar a partir de la base de datos de referencia. Se ha obtenido la media de los resultados para evaluar las prestaciones de la detección y se ha probado finalmente el sistema con un conjunto de test grabado en condiciones reales.



a) Matriz de confusión con ruido de fondo variable (música)



b) Matriz de confusión con ruido de fondo contínuo (aspiradora).

Figura 5. Matrices de confusión para dos realizaciones de prueba sobre el data set de referencia.

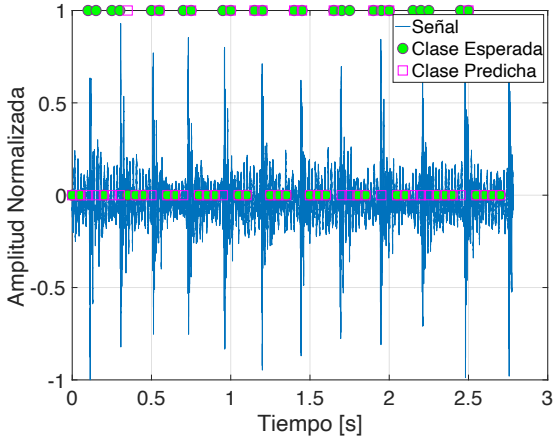


Figura 6. Detalle del resultado de la detección de gotas. S/N =10 dB.

La figura 5 muestra dos ejemplos de ejecución del clasificador, con una duración de trama de 50 ms:

- La figura 5a) muestra los resultados obtenidos para un goteo con ruido de fondo variable. De las 600 tramas del conjunto de prueba, se detectan adecuadamente un 59% de las tramas con gotas, presentando un 0,6 % de falsos positivos, que se corresponden con pasajes impulsivos dentro del fondo musical.
- La figura 5b) muestra el caso de funcionamiento con ruido estacionario. Se observa una tasa de detección de gotas del 35 % y no se presentan falsos positivos.

La figura 6 muestra en detalle cómo funciona la clasificación sobre una señal real captada con ruido de fondo de un goteo percutiendo sobre un fregadero de aluminio. Cada gota dura aproximadamente 4 tramas, y en prácticamente todas las gotas se detecta adecuadamente una de las cuatro tramas como gotas (esencialmente debido a la influencia del ruido de fondo). Postprocesando los resultados, tomando la decisión cada 4 tramas:

1. Si no hay detección positiva en alguna de las 4 tramas, se decide que no hay gota
2. Si al menos una de las 4 tramas presenta resultado positivo, se decide que existe un goteo.

De esta forma, siempre que la tasa de detección de tramas pertenecientes a la clase "*drop*" sea superior a un 25 %, el sistema detectará con garantías la presencia de un goteo.

5. CONCLUSIONES

En este trabajo se ha evaluado la posibilidad de detectar la aparición de goteos de agua en entornos domésticos en condiciones de ruido variable. Se generó una base de datos mínima con sonidos reales, que se expandió con señales obtenidas de modelos sintéticos. Se generaron conjuntos de entrenamiento y prueba utilizando la kurtosis, la función de

densidad de probabilidad de las altas frecuencias (f>2000 Hz) y la velocidad de variación de los MFCC para alimentar un clasificador SVC. La tabla 1 muestra los resultados medios obtenidos a partir de 10 pruebas con distintas combinaciones de conjuntos de prueba y test.

	Continuous Noise	Background Music	Real Test
Accuracy (%)	93.55	90.33	72.73
Recall (%)	59.55	41.18	44.00
Specificity (%)	99.41	98.45	96.67
Precision(%)	94.64	81.40	91.67
FMeasure	0.73	0.63	59.46

Tabla 1: Medida de las prestaciones del clasificador.

Los parámetros para la evalación de la calidad del clasificador son:

- *Accuracy*. Mide el porcentaje de tramas correctamente clasificadas sobre el total de predicciones (verdaderos positivos+falsos negativos).
- *Recall*. Mide el porcentaje de eventos correctamente clasificados (verdaderos positivos) sobre el total de predicciones.
- *Precision.* Mide la fracción de verdaderos positivos sobre el total de positivos esperados.
- F-Score. Es una media armónica de la precisión y el recall:

$$FScore = \frac{2PrecisionxRecall}{Precision + REcall} \qquad (8)$$

Los resultados presentan una alta precisión y un bajo Recall lo que quiere decir que aunque un número importante de tramas que pertenecen a la clase "*drop*" son clasificadas incorrectamente, las detectadas son correctas con una alta probabilidad. La probabilidad de falsa alarma, por tanto, es muy baja. Los test realizados muestran que es posible implementar un detector de goteos a partir de la monitorización acústica.

12. REFERENCIAS

[1] Rosa M. Alsina Pagès, Joan Navarro, Francesc Alías, and Marcos Hervás. homesound: Real-time audio event detection based on high performance computing for behaviour and surveillance remote monitoring. Sensors (Basel), 17, April 2017.

[2] Shivan Sharma et al. An intelligent system for infant cry detection and information in real time. In Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2017. doi: 10.1109/ACIIW. 2017.8272600.

[3] Shubham Asthana, Naman Varma, and Vinay Mittal. Preliminary analysis of causes of infant cry. In IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pages 000468–000473, 12 2014. doi: 10.1109/ ISSPIT.2014.7300634.

[4] Jiaxing Ye, Takumi Kobayashi, and Masahiro Murakawa. Urban sound event classification based on local and global features aggregation. Applied Acoustics, 2017. doi: https://doi.org/10.1016/j.apacoust.2016.08.002.

[5] Dan Ellis Tuomas Virtanen, Mark D. Plumbey, editor. Computational Analysis of Sound Scenes and Events. Springer, 2018.

[6] BBC sound effects. URL http://bbcsfx.acropolis.org.uk/.

[7] M. Kob and M. Vörlander. Band filters and short reverberation times. Acustica united with Acta Acustica, 86:350–357, 2000.

[8] Patrice Guyot. Julien Pinquier, Régine André-Obre. Water sound recognition based on physical models. In IEEE International Conference on Acoustics, Speech, and Signal Processing – ICASSP 2013, pages 793–797, Vancouver, May 2003.

[9] William Moss and Hengchin Yeh. Sounding liquids:automatic sound synthesis from fluid simulation. In ACM Transactions on Graphics, volume 28, December 2009.

[10] Iryna Skrypnyk. Irrelevant features, class separability, and complexity of classification problems. In IEEE 23rd International Conference on Tools with Artificial Intelligence, Boca Raton, Fl. USA, November 2011.

[11] Iman Khosravi, Abdolreza Safari, and Saeid Homayouni. Msmd: maximum separability and minimum dependency feature selection for cropland classification from optical and radar data. International Journal of Remote Sensing, 39(8):2159– 2176, 2018. doi: 10.1080/01431161.2018.1425564.

[12] Xiaodan Zhuang et al. Feature analysis and selection for acoustic event detection. In IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas,NV, USA, April 2008.

[13] Eva Vozáriková, Jozef Juhár, and Anton Cižmár. Acoustic events detection using MFCC and MPEG-7 descriptors. In Czyzewski A. Dziech A., editor, Multimedia Communications, Services and Security. MCSS 2011. Communications in Computer and Information Science,, volume 149. Springer, Berlin, Heidelberg, 2011.

[14] Courtenay V. Cotton and Daniel P. W. Ellis. Spectral vs. spectro-temporal features for acoustic event detection. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, November 2011.

[15] Cédric Gervaise, A Barazzutti, Sylvain Busson, Y Simard, and N Roy. Automatic detection of bioacoustics impulses based on kurtosis under weak signal to noise ratio. Applied Acoustics, 71:1020–1026, 11 2010. doi: 10.1016/j.apacoust.2010.05.009.

[16] J.A.K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. Neural Processing Letters, June 1999.

[17] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. IEEE TRANSACTIONS ON NEURAL NETWORKS, 13(2), march 2002.

[18] T. G. Leighton. The acoustic buble. Academic Press, 1994.

[19] B. W. Shuller. Intelligent Audio Analysis. Springer, 2013