ISSN 1870-4069

Clasificación contextual de vocalizaciones de perros de asistencia apoyada por segmentación automática

Roilhi Frajo Ibarra-Hernández[1], Luis Villaseñor-Pineda[2], Humberto Pérez-Espinoza[3], Hugo Jair Escalante[2]

[1] Instituto Nacional de Astrofísica, Óptica y Electrónica, Consejo Nacional de Ciencia y Tecnología, México

[2] Instituto Nacional de Astrofísica, Óptica y Electrónica, Coordinación de Ciencias Computacionales, México

[3] Centro de Investigación Científica y de Educación Superior de Ensenada, Unidad de Transferencia Tecnológica, México

roilhi@inaoe.mx, {villasen, hugojair}@inaoep.mx, hperez@cicese.mx,

Resumen. El Aprendizaje Automático ha permitido el desarrollo de métodos de detección de emociones a partir de las vocalizaciones humanas. Este concepto se ha extrapolado al manejo de contextos que describan el estado emocional o fisiológico de los perros de búsqueda y asistencia, con el propósito de predecir y controlar su comportamiento. Sin embargo, para el procesamiento de la señal de audio es importante considerar únicamente aquellas partes que contienen información útil para mejorar el rendimiento del clasificador, esta etapa es conocida como segmentación. En este trabajo se compara el desempeño del algoritmo de clasificación de contextos en las vocalizaciones de un conjunto de perros asistido por segmentación automática, con lo cual es evidente el ahorro en recursos y tiempo de ejecución. Los resultados obtenidos muestran una mejora en la métrica F1-score con respecto a la clasificación asistida por segmentación manual, además de un buen desempeño en la detección de ladridos, dados los altos niveles de relación señal a ruido (SNR) obtenidos entre los audios de la base de datos empleada.

Palabras clave: Clasificación de ladridos, segmentación automática, aprendizaje automático, descriptores acústicos.

Assistance Dogs Bark Classification Supported by Automatic Segmentation

Abstract. Machine Learning has allowed the development of methods to detect emotions from human vocalizations. This concepts has been taken to handle

the context which describe the emotional or physiological states in assistance dogs, when having as a goal to predict and control its behavior. However, to process the audio signal it is relevant to consider those parts which contain useful information to enhance the classification performance, this stage is known as segmentation. In this work, we compare the classifier's algorithm efficiency when it is assisted by automatic segmentation, which also improves the execution time and resources. The obtained results show an increase in the F1-score with respect to the classification scheme assisted by manual segmentation, as well as accuracy when detecting barks, given by the high levels of signal-to-noise ratio (SNR) obtained among the database used.

Keywords: Bark classification, automatic segmentaetion, machine learning, acoustic descriptors.

1. Introducción

La comunicación es una herramienta fundamental en las interacciones sociales para transmitir ideas, pensamientos y estados afectivos. Para expresar algún estado emocional interno, los seres humanos y animales usan diversas expresiones comunicativas, entre las que se encuentran las expresiones vocales.

Dentro del conjunto de vocalizaciones humanas, se sigue un conjunto de reglas simples para codificar el estado interno del hablante a través de parámetros acústicos. La especie humana ha sido capaz de usar estas reglas estructurales para asociar ladridos de perros con diversos contextos que representen su estado emocional [4].

En efecto, gracias a la domesticación ha sido posible que los perros hayan desarrollado habilidades sociales y de comunicación [14]. Estas habilidades han conducido al involucrar a los perros en las tareas de búsqueda y asistencia, en las cuales se desarrolla un proceso de entrenamiento para apoyar a las tareas de rescate y apoyo a personas con alguna discapacidad física o motriz.

Por este motivo, se ha consolidado como un tema importante el conocer el estado interno y emocional de los perros entrenados a través de los contextos de sus vocalizaciones. Por medio de un contexto definido en su ladrido se describen diversos estados psico-emocionales que permitan comprender directamente alguna acción del perro: hambre, soledad, enojo, felicidad, etc.

La comunicación interactiva canino-humana ha sido objeto de estudio debido a que aún existen diversas preguntas abiertas sobre el significado contextual de las vocalizaciones.

Particularmente los etólogos, quienes se dedican a estudiar el comportamiento de los animales, tienen el interés de crear perfiles de conducta para monitorear y predecir el comportamiento de los perros domésticos, especialmente aquellos dedicados al rescate, lazarillos entre otros que tengan algún entrenamiento especializado.

Existen caracterísitcas acústicas importantes en los ladridos y vocalizaciones del perro doméstico, tales como la frecuencia, amplitud, tono, ritmo entre otras [12]. Dichos parámetros son posibles de relacionarse con algún estado emocional, actitud, reacción fisiológica o estado particular del perro, el cual se conoce como *contexto* [6].

Tabla 1. Distribución de etiquetas (contextos) en las señales de audio de vocalizaciones de la base de datos Mudi [13].

contexto	**muestras**
ball	53
stranger	46
food	41
fight	30
walk	29
play	23
alone	22

El presente trabajo tiene como propósito fortalecer una etapa importante en la clasificación de contextos en las vocalizaciones de perros de asistencia: la segmentación. Se ha motivado por los resultados de otras investigaciones previas donde se han seleccionado como características de clasificación descriptores acústicos de bajo (LLDs) y alto nivel (HLDs) [9, 11] para alimentar un clasificador de Aprendizaje Automático basado en Máquinas de Soporte Vectorial (SVM) [2].

Entre los parámetros de caracterización de las vocalizaciones caninas, se encuentran los coeficentes cepstrales de frecuencia Mel (MFCCs), ampliamente empleados por su eficiencia en la detección y clasificación de voz, así como los espectrograms Mel (Melspec) además de otros parámetros espectrales, tales como la energía, flujo espectral, centroide, desvanecimiento, entre otros.

No obstante, en dichas investigaciones la clasificación realizada requiere incluir el proceso de segmentación, el cual consiste en delimitar los tiempos de inicio (*onset*) y fin (*offset*) de ladrido, con el objetivo de extraer y procesar las muestras útiles de la señal y descartar las pausas entre dichos eventos. Para la ejecución de la segmentación, Pérez et al., [9] ha realizado un procesamiento manual, requiriendo el uso de *anotadores* auxiliares en la tarea de registrar los tiempos de onset y offset para cada señal.

Esto puede ser un proceso exhaustivo y monótono para ser desarrollado por humanos, siendo suceptible a errores. Para evitar este consumo de recursos humanos y tiempo, en este trabajo de investigación se propone asistir la clasificación por medio de mecanismos de segmentación automática.

En este trabajo se ha seleccionado un método de segmentación basado en la detección de tosidos en señales acústicas respiratorias [7], debido a que estas señales tienen similitudes con las vocalizaciones caninas en sus componentes espectrales.

Para evaluar si es adecuado el complementar la clasificación conn el método automático de segmentación, se comparará el desempeño del algoritmo por medio de la métrica F1-score con y sin segmentación.

El artículo se divide de la siguiente manera: en la sección 2 se describirá la base de datos utilizada y los contextos objetivo de la clasificación, así como una descripción detallada del algoritmo de segmentación automática, además de la extracción y selección de características.

Posteriormente, en la sección 3 se describen los procedimientos, parámetros y algoritmos de clasificación empleados, así como los resultados producidos de la experimentación.

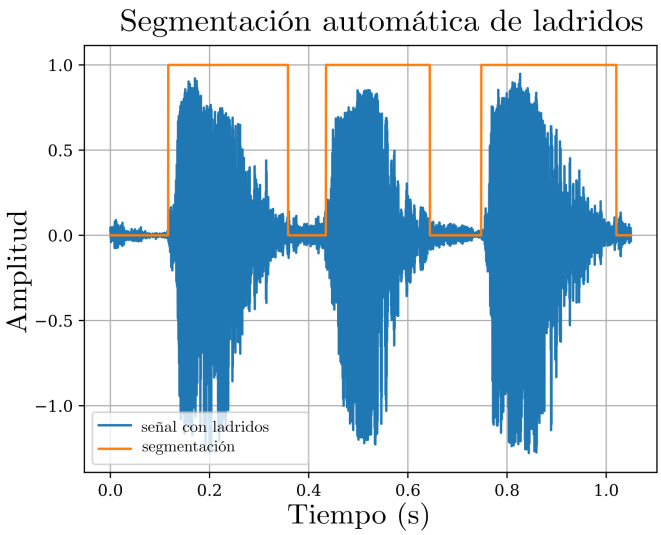

Fig. 1. Realización de la segmentación del audio con vocalizaciones. Se han detectado 3 eventos de ladrido automáticamente.

Finalmente, la sección 4 expone las conclusiones referentes a los resultados y hallazgos derivados de los experimentos conducidos en este trabajo, además de las líneas a seguir de trabajo futuro.

2. Metodología

2.1. Conjunto de datos de vocalizaciones

Para el desarrollo de los experimentos realizados en esta investigación, se ha empleado el conjunto de datos de Póngracz et al. [13], el cual comprende señales de audio de ladridos a través de los perros de la raza *Mudi*, cuyo origen proviene de Hungría. Las vocalizaciones han sido etiquetadas en siete diferentes contextos, que se muestran a continuación:

1. **Alone:** Se aisló al perro en un área exterior, atándolo. El dueño caminó lejos de la vista del perro.
2. **Ball:** El dueño sostenía una pelota o juguete a 1.5 m enfrente del perro.
3. **Fight:** El entrenador atacó al dueño y al perro. El dueño mantuvo al perro con correa.
4. **Food:** El dueño sostenía un platón de comida a 1.5 m enfrente del perro.
5. **Play:** El dueño jugó algún juego con el perro.
6. **Stranger:** El experimentador apareció en el terreno donde el perro suele convivir o enfrente del perro.
7. **Walk:** El dueño simuló una serie de acciones tal como si fuese a sacar al perro a pasear.

Tabla 2. Distribución de muestras por cada contexto después de realizar la segmentación automática.

contexto	muestras
alone	826
ball	1025
fight	1008
food	829
play	735
stranger	1318
walk	916

Los ladridos fueron registrados en un diferente número de sesiones para cada perro. Las locaciones en que fueron grabados los audios fueron las residencias de los dueños, a excepción de los contextos *Alone* y *Fight*. Se digitalizaron las grabaciones a una frecuencia de muestreo de 22.05 kHz y 16 bits por muestra. Originalmente, éstas fueron registradas con una grabadora de cinta y un micrófono. Se reescaló la amplitud de las formas de onda tal que el pico se ubicase en -6 dB.

La base de datos contiene un total de 244 registros de audio con un formato .wav, cuyas longitudes en tiempo se encuentran en el intervalo $[1.03, 378.24]$ segundos. La Tabla 1 muestra la cantidad de audios etiquetados para cada uno de los contextos de la base Mudi. Se observa que la clase mayoritaria (con mayor número de muestras) está dada por el contexto *ball*, mientras que la clase minoritaria se presenta en el contexto *alone*.

2.2. Segmentación automática

Al igual que otras señales de audio, los registros de vocalizaciones presentan partes con silencios, generadas por las pausas entre ladridos. Sin embargo, con el objetivo de discriminar estas partes y conservar los segmentos con ladridos como unidades de análisis, se ha empleado un método de segmentación automática basado en señales de tosidos humanos [7], dado que esta señal presenta similitudes en cuanto a la duración temporal y contenido frecuencial con los ladridos [1].

Para efectos de incrementar la rapidez en el cálculo, el audio del ladrido se submuestreó a 8kHz, permitiendo una visualización frecuencial de las componentes frecuenciales adecuadas para poder detectar las características de la señal [12].

El algoritmo de segmentación está basado en el cálculo de la relación señal a ruido SNR, se compara mediante una *histéresis* las regiones de la forma de onda cuyos picos sufren cambios rápidos en potencia.

La señal deberá ser normalizada en el intervalo $[-1, 1]$, para identificar más rápidamente estos cambios. El cálculo de la SNR se realiza mediante el procedimiento descrito en la Ecuación 1 :

$$SNR = 20 \log_{10} \left(\frac{\sqrt{\frac{1}{|x_s|} \sum_{x(n) \in x_s} x(n)^2}}{\sqrt{\frac{1}{|x_n|} \sum_{x(n) \in x_n} x(n)^2}} \right), \tag{1}$$

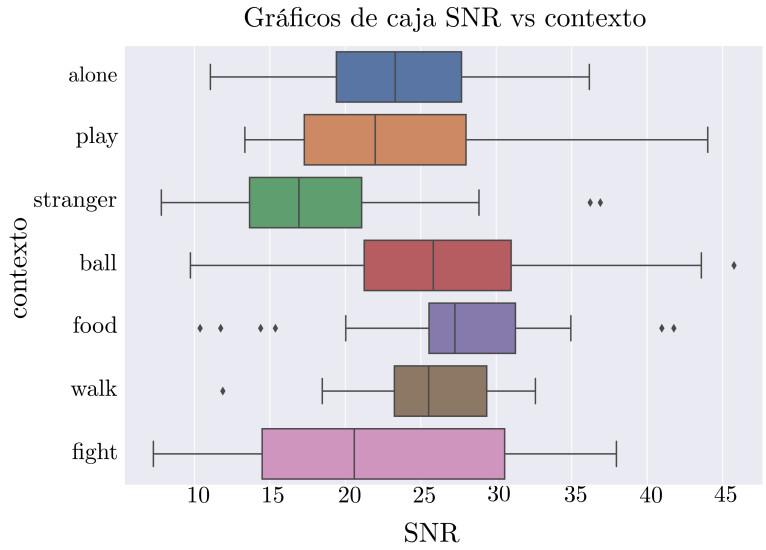

Fig. 2. Gráficos de caja para mostrar la distribución de la SNR por contexto.

donde x_s es una muestra de señal (tosido/ladrido) determinada durante el proceso de segmentación, x_n una muestra de ruido (muestra que no es considerada como señal) y $x(n)$ es la secuencia completa de audio que contiene muestras tanto del tipo x_s como del tipo x_n.

En concreto, la SNR nos indicará el cociente de la energía de los eventos de ladrido y las pausas, teniendo en esta magnitud una referencia para evaluar el desempeño del segmentador y su capacidad de detección de ladridos.

La Figura 1 muestra la segmentación automática aplicada a un segmento de audio que contiene 3 ladridos. Se ha observado que el algoritmo es capaz de detectar diversos tipos de ladridos. Sin embargo, existen otros parámetros necesarios de ajustar para poder obtener resultados más precisos, tales como:

- *padding:* Intervalo de tiempo mínimo entre eventos. Es decir, la longitud de tiempo mínima al inicio y al final de un ladrido. En este caso de acuerdo con lo realizado por Póngracz et al., [12] se estableció en 5 ms (*inter-bark interval*).
- *min_length:* Intervalo de tiempo mínimo a considerar como un evento de ladrido. Es decir, la longitud mínima en tiempo que pudiese tener un ladrido. Se estableció este parámetro en 20 ms.
- *th_l_multiplier:* Umbral de energía mínimo para el comparador de histéresis, en valor de raíz cuadrática media (RMS). Se estableció como 0.2.
- *th_h_multiplier:* Umbral de energía máximo para el comparador de histéresis, en valor de raíz cuadrática media (RMS). Se estableció como 2.

Derivado del proceso de segmentación, se han detectado 6657 eventos de ladrido. La distribución de muestras para cada clase se muestra en la Tabla 2, donde ahora el contexto *stranger* representa la clase mayoritaria, mientras que el contexto *play* la clase minoritaria.

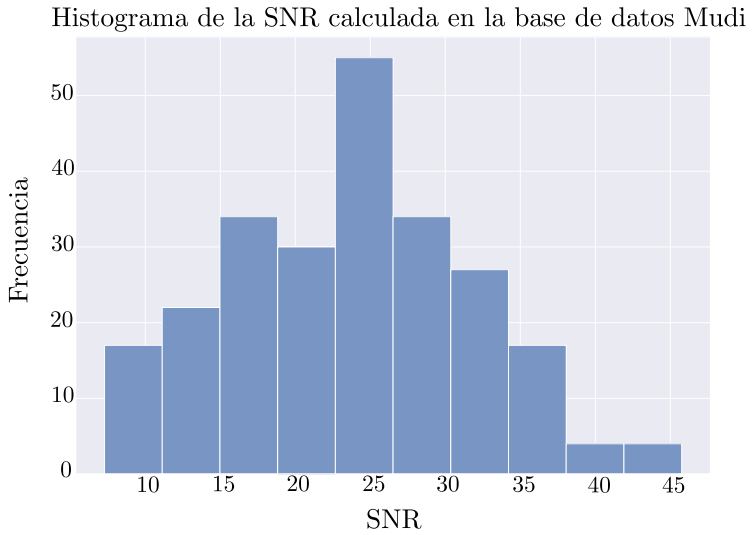

Fig. 3. Histograma general del cálculo de la SNR a cada señal de la base Mudi.

Para verificar el desempeño del algoritmo de segmentación, se calcula la SNR de acuerdo con la ecuación 1, donde se pondera la trama que ha sido detectada como ladrido frente a las pausas o silencios, que son considerados como *ruido*.

La Figura 3 muestra un histograma del cálculo de la SNR a cada audio de la base Mudi. Se observa que el valor de la media se encuentra entre los 15 y 30 dB, además, en cada una de las muestras existieron segmentos de ladrido correctamente detectados, ya existieron cerca de 20 señales con una SNR cercana a 10 dB.

La verificación del desempeño por medio de la SNR también fue distribuída de acuerdo con los contextos y nombres de los individuos de la base de datos Mudi. En la Figura 2 se muestra la distribución de la SNR por contexto, donde se corrobora que el valor de la mediana está localizado entre los 15 y 30 dB, tal y como ocurrió en el histograma general.

De manera particular, los valores más bajos de SNR fueron calculados en el contexto *stranger* y *fight*, esto se debe a los pequeños tiempos de intervalo entre ladridos. De igual manera se observa una más alta variabilidad en la SNR del contexto *fight*, lo cual es también proporcional a la variación alta en las longitudes de semento detectadas para este contexto.

Por otra parte, la Figura 4 muestra la distribución de SNR por cada uno de los 12 individuos. Particularmente se observa una variabilidad alta en la mayoría de los perros, excepto romanecsutka y romannefecske.

2.3. Extracción de características

Las características extraídas para la experimentación de este trabajo están basadas en descriptores de bajo nivel (LLDs), lo cual fue realizado mediante la herramienta openSMILE [3].

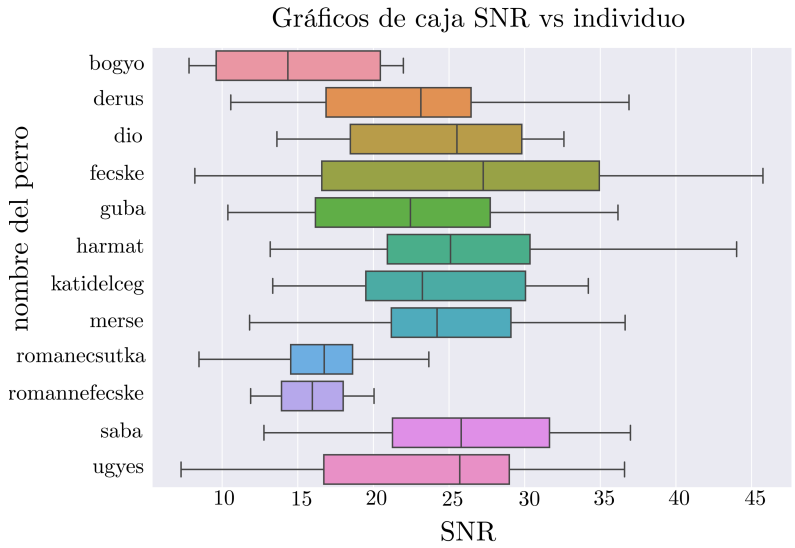

Fig. 4. Gráficos de caja para mostrar la distribución de la SNR por individuo.

Este software presenta tiempos de cómputo muy ágiles y además puede integrarse de manera muy sencilla con el lenguaje de programación Python, el cual ha sido empleado para realizar los experimentos en esta investigación, debido a la popularidad de las librerías integradas para aprendizaje automático, tales como scikit-learn.

Las características seleccionadas están basadas en el conjunto *emolarge*, diseñado para la detección de emociones en la voz humana y además previamente utilizado para la clasificación de vocalizaciones de ladridos de perros por medio de contextos con segmentación de tipo manual [9].

2.4. Selección de características

Por medio del conjunto *emolarge* de openSMILE es posible extraer 6652 parámetros LLDs, usando una ventana de 25 ms y un corrimiento de 10 ms, además del cálculo de coeficientes de regresión delta y doble-delta.

Con el objetivo de establecer una comparación entre el trabajo previo de clasificación con segmentación manual [9] y la segmentación automática propuesta, se ha seleccionado el método *Relief Attribute* como técnica de selección de características, para agilizar el entrenamiento y reducir la dimensionalidad de dicho conjunto. La herramienta Weka [5] se ha empleado para llevar a cabo esta tarea.

En la Tabla 3 se muestra la cantidad de parámetros originales que extrae el conjunto *emolarge* de la librería openSMILE, la cantidad de características reducidas por Pérez et al. [9] haciendo uso de la segmentación manual y finalmente el número de parámetros que se han reducido por medio del método *Relief Attribute* en weka añadiendo el método de segmentación automática propuesto en el presente trabajo. Se observa que los parámetros mayormente seleccionados forman parte de las categorías **mfcc** y **melspec** respectivamente.

Tabla 3. Comparación de las cantidades de parámetros del conjunto *emolarge* de openSMILE, la clasificación realizada con segmentación manual [9] y con segmentación automática (reportada en este trabajo) usando una reducción *Relief*.

Características	**Cantidad de parámetros**		
	emolarge	**Pérez et al.** [9]	**este trabajo**
mfcc	1521	224	196
melspec	3042	66	201
energy	117	37	12
fband	468	8	9
RollOff	468	69	29
Flux	117	4	25
spectralCentroid	117	17	7
MinPos	117	0	3
voiceProb	117	26	15
F0	117	20	3

3. Clasificación de los ladridos

3.1. Métodos de la clasificación

Para conducir los experimentos de la clasificación de contextos a partir de los datos extraídos, se ha empleado el algoritmo de máquinas de soporte vectorial (SVM), el cual ha demostrado buenos resultados para dicha tarea en trabajos previos [9, 10].

Se seleccionó un kernel de tipo *polinomial* para el método de SVM. El conjunto de datos se ha dividido en dos subconjuntos aleatorios de *entrenamiento* y *prueba* considerando el 80 % y el 20 % respectivamente.

Para evitar el sobre-ajuste del modelo, se empleó la técnica de validación cruzada (*CV*), agregando métodos de *estratificación* para asegurar que en cada partición se empleará un número de muestras balanceado por cada clase.

3.2. Calibración de hiperparámetros

Dado que la distribución de características muestra clases que no son linealmente separables y el kernel es de tipo polinomial, la SVM empleada en el algoritmo de clasificación consiste en un híperplano de máximo margen.

Dentro de los hiperparámetros a considerar en este método se encuentran el grado del polinomio y el parámetro C, considerado como un *hiperparámetro de calibración*. A través de C el discriminador controla la cantidad de violaciones que pueden darse en los márgenes del híperplano, así como la severidad de las mismas.

En resumen, mediante C se controla el balance entre el *sesgo* y la varianza del modelo. Para evaluar algunas combinaciones posibles entre el grado del polinomio y la calibración C se ha empleado el método de la búsqueda por cuadrícula mediante validación cruzada (*Grid-Search CV*) por sus siglas en Inglés.

Se han evaluado la métrica *F1-score* y un número de particiones o *folds* $N_{folds} = 10$. Se probaron las combinaciones entre los parámetros $C = [1, 10, 100, 100]$ y $p = [2, 3, 4, 5, 6]$ donde p es el grado del polinomio.

Tabla 4. Comparación de los resultados obtenidos para los experimentos del agrupamiento de contextos.

Núm. de experimento	Pérez et al. [9]	Este trabajo
1	0.85	0.88
2	0.85	0.88
3	0.78	0.82

La mejor combinación de parámetros resultó al seleccionar $p = 3$ y $C = 100$ con una puntuación f1-score promedio de $\mu_{f1} = 0.7435$ y desviación estándar $\sigma_{f1} = 0.0178$. La experimentación se condujo en la herramienta scikit-learn [8].

3.3. Agrupamiento de contextos

Se han definido agrupaciones de contextos de acuerdo con la *valencia* y *excitación* de cada uno de ellos. Estos modelos han sido frecuentemente usados para la detección de emociones humanas [15].

Se conoce como Valencia a la atracción (positiva) o aversión (negativa) intrínseca de cada contexto, mientras que la excitación corresponde al nivel reactivo o sensible del contexto. Con base en el trabajo realizado por Pérez et al. [9], los contextos han sido agrupados de la siguiente manera:

- **Experimento 1:** Valencia Negativa (Fight, Stranger, Alone) vs Valencia Positiva (Walk, Ball, Play, Food).
- **Experimento 2:** Alta Excitación (Fight, Stranger, Wall, Ball, Play) vs Baja Excitación (Alone, Food).
- **Experimento 3:** Valencia Negativa y Alta Excitación (Fight, Stranger) vs Valencia Positiva y Alta Excitación (Walk, Ball, Play) vs Baja Excitación(Alone, Food).

En la Tabla 4 se muestran los resultados producidos tras conducir los experimentos de agrupación de contextos. Se ha observado un aumento en la métrica f1-score en todos los experimentos al introducir en este trabajo la segmentación automática frente a la segmentación manual del algoritmo desarrollado por Pérez et al. [9].

3.4. Clasificación individual de los contextos

Como experimento final, se probó el desempeño de la clasificación con el subconjunto de prueba, correspondiente a una selección aleatoria proporcional al 20 % del conjunto de datos. De igual manera, se consideró la métrica F1, en conjunto con la precisión y el *recall*. Los resultados obtenidos se muestran en la Tabla 5.

4. Conclusiones

En el presente trabajo se ha evaluado el desempeño de la clasificación de los contextos emocionales en los ladridos al introducir un algoritmo de segmentación automática.

Tabla 5. Resultados de la clasificación de contextos en ladridos sobre el subconjunto de prueba.

class	**precision**	**recall**	**f1-score**
alone	0.82	0.76	0.79
ball	0.62	0.65	0.64
fight	0.87	0.86	0.86
food	0.70	0.79	0.74
play	0.71	0.65	0.68
stranger	0.78	0.82	0.80
walk	0.73	0.65	0.69
precisión total	0.75		
macro avg	0.75	0.74	0.74
media ponderada	0.75	0.75	0.75

Dicho algoritmo está basado en la detección de señales auditivas correspondientes a tosidos, cuyo contenido frecuencial y duración en tiempo presentan similitudes con respecto a los ladridos.

Los resultados muestran una mejora en el desempeño, al incrementar la métrica F1-score al hacer una agrupación de contextos de acuerdo a la valencia y a la excitación. Posteriormente se ha determinado que la clasificación presenta los mejores resultados al calibrar hiperparámetros mediante una búsqueda en cuadrícula de la validación cruzada.

Para el clasificador seleccionado, dado por un algoritmo de máquinas de soporte vectorial con kernel polinomial, se ha demostrado que el orden $p = 3$ y el hiperparámetro de calibración $C = 100$ son óptimos para la clasificación, ya que mediante estos se produjeron las métricas f1 más altas.

Se ha comparado el desempeño de la clasificación al introducir la técnica *Relief* para la selección de características. De igual manera, se ha mejorado el desempeño al realizar un agrupamiento de contextos en valencia positiva y negativa, así como excitación alta y baja.

Los resultados muestran un incremento en la métrica F1 de 0.85 a 0.88 al clasificar mediante la valencia, del mismo modo en el agrupamiento mediante excitaciones. Para la tercer experimentación, conjuntando valencia con excitación alta, el F1 ha incrementado de 0.78 a 0.82.

En general, se ha encontrado que los contextos de valencia alta: *alone*, *fight* y *stranger* son más aptos para clasificar mediante los métodos presentados en este trabajo, ya que presentaron métricas F1 de 0.80, 0.84 y 0.79 respectivamente. De igual manera, se ha comprobado la robustez del algoritmo de segmentación automática aplicado, ya que los valores medios de SNR oscilan entre los 15 y 30 dB, teniendo con ello una alta precisión en discriminar las pausas y detectar eficientemente los eventos de ladrido.

Finalmente, se ha evaluado el desempeño del algoritmo al clasificar los contextos del subconjunto de entrenamiento, obteniendo una precisión ponderada de 0.75, resultado cercano a lo obtenido por Perez et al., [9] usando segmentación manual.

De manera similar, se ha comprobado que los contextos de alta valencia como *fight* y *alone* presentan las métricas F1 más altas y por lo tanto mayor potencial de ser detectados mediante los métodos expuestos en este trabajo.

Dados los resultados del presente trabajo mediante diversas experimentaciones, se ha demostrado que la segmentación automática no sólo es una herramienta que contribuye a automatizar el proceso de clasificación emocional de ladridos y eficientar el tiempo, sino que además mejora el desempeño del clasificador.

Como trabajo futuro se aplicará el método de segmentación automática a bases de datos más extensas, con el objetivo de automatizar el proceso y evitar los cuellos de botella en el mismo. Además, se utilizarán técnicas de aprendizaje profundo.

Agradecimientos. El primer autor agradece al proyecto *TZUKU: Métodos computacionales para el análisis del comportamiento de perros de búsqueda y asistencia* por su apoyo en el financiamiento de este trabajo de investigación. También se agradece al Laboratorio de Súper Cómputo del INAOE por facilitar sus recursos para el desarrollo de los experimentos conducidos en este trabajo de investigación, al igual que al Consejo Nacional de Ciencia y Tecnología (CONACYT).

Referencias

1. Chang, A. B.: The physiology of cough. Paediatric Respiratory Reviews, vol. 7, no. 1, pp. 2–8 (2006) doi: 10.1016/j.prrv.2005.11.009
2. Cortes, C., Vapnik, V.: Support-vector network. Machine Learning, vol. 20, pp. 273–297 (1995)
3. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: The munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM International Conference on Multimedia, pp. 1459–1462 (2010) doi: 10.1145/1873951.1874246
4. Faragó, T., Takács, N., Miklósi, Á., Pongrácz, P.: Dog growls express various contextual and affective content for human listeners. Royal Society Open Science, vol. 4, no. 170134, pp. 1–11 (2017) doi: 10.1098/rsos.170134
5. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H.: The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10–18 (2009) doi: 10.1145/1656274.1656278
6. Molnár, C., Kaplan, F., Roy, P., Pachet, F., Pongrácz, P., Dóka, A., Miklósi, Á.: Classification of dog barks: A machine learning approach. Animal Cognition, vol. 11, pp. 389–400 (2008) doi: 10.1007/s10071-007-0129-9
7. Orlandic, L., Teijeiro, T., Atienza, D.: The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. Scientific Data, vol. 8, no. 156, pp. 1–10 (2021) doi: 10.1038/s41597-021-00937-4
8. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in python. The Journal of Machine Learning Research, vol. 12, pp. 2825–2830 (2011)
9. Pérez-Espinosa, H., Pérez-Martınez, J. M., Durán-Reynoso, J. Á., Reyes-Meza, V.: Automatic classification of context in induced barking. Research in Computing Science, vol. 100, pp. 63–74 (2015) doi: 10.13053/rcs-100-1-6
10. Pérez-Espinosa, H., Reyes-García, C. A., Villaseñor-Pineda, L.: Acoustic feature selection and classification of emotions in speech using a 3D continuous emotion model. Biomedical Signal Processing and Control, vol. 7, no. 1, pp. 79–87 (2012) doi: 10.1016/j.bspc.2011.02.008

11. Pérez-Espinosa, H., Reyes-Meza, V., Aguilar-Benitez, E., Sanzón-Rosas, Y. M.: Automatic individual dog recognition based on the acoustic properties of its barks. Journal of Intelligent & Fuzzy Systems, vol. 34, no. 5, pp. 3273–3280 (2018) doi: 10.3233/JIFS-169509
12. Pongrácz, P., Molnár, C., Miklósi, Á.: Acoustic parameters of dog barks carry emotional information for humans. Applied Animal Behaviour Science, vol. 100, no. 3-4, pp. 228–240 (2006) doi: 10.1016/j.applanim.2005.12.004
13. Pongrácz, P., Molnár, C., Miklósi, A., Csányi, V.: Human listeners are able to classify dog (canis familiaris) barks recorded in different situations. Journal of Comparative Psychology, vol. 119, no. 2, pp. 136–144 (2005) doi: 10.1037/0735-7036.119.2.136
14. Range, F., Virányi, Z.: Tracking the evolutionary origins of dog-human cooperation: The "Canine Cooperation Hypothesis". Frontiers in psychology, vol. 5, no. 1582, pp. 1–10 (2015) doi: 10.3389/fpsyg.2014.01582
15. Scherer, K. R.: Psychological models of emotion. The neuropsychology of emotion, pp. 137–162 (2000)