

UNIVERSIDAD NACIONAL DEL LITORAL



DOCTORADO EN INGENIERÍA

# **Reconocimiento de actividades a partir de señales inerciales y acústicas**

Sebastián Rodrigo Vanrell

**FICH**

FACULTAD DE INGENIERÍA  
Y CIENCIAS HÍDRICAS

**INTEC**

INSTITUTO DE DESARROLLO TECNOLÓGICO  
PARA LA INDUSTRIA QUÍMICA

Tesis de Doctorado **2018**





UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Ingeniería y Ciencias Hídricas  
Instituto de Desarrollo Tecnológico para la Industria Química

# **RECONOCIMIENTO DE ACTIVIDADES A PARTIR DE SEÑALES INERCIALES Y ACÚSTICAS**

**Sebastián Rodrigo Vanrell**

Tesis remitida al Comité Académico del Doctorado  
como parte de los requisitos para la obtención  
del grado de  
DOCTOR EN INGENIERIA  
Mención Inteligencia Computacional, Señales y Sistemas  
de la  
UNIVERSIDAD NACIONAL DEL LITORAL

**2018**

Comisión de Posgrado, Facultad de Ingeniería y Ciencias Hídricas, Ciudad Universitaria, Paraje "El Pozo",  
S3000, Santa Fe, Argentina.





UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Ingeniería y Ciencias Hídricas  
Instituto de Desarrollo Tecnológico para la Industria Química

## RECONOCIMIENTO DE ACTIVIDADES A PARTIR DE SEÑALES INERCIALES Y ACÚSTICAS

**Sebastián Rodrigo Vanrell**

**Lugar de Trabajo:**

Instituto de Investigación en Señales, Sistemas e Inteligencia  
Computacional –  $\text{sinc}(i)$ , FICH-UNL/CONICET

**Director:**

Dr. H. Leonardo Rufiner       $\text{sinc}(i)$ , FICH-UNL/CONICET  
FI-UNER

**Co-director:**

Dr. Diego H. Milone       $\text{sinc}(i)$ , FICH-UNL/CONICET

**Jurado Evaluador:**

Dr. Omar Chiotti      INGAR, CONICET-UTN  
Dr. Pablo Mandolesi      IIIIE, CONICET-UNS  
Dr. Humberto Torres      INIGEM, CONICET-UBA  
Dr. Mariano Rubiolo       $\text{sinc}(i)$ , FICH-UNL/CONICET  
UTN-FRSF

**2018**





UNIVERSIDAD NACIONAL DEL LITORAL  
Facultad de Ingeniería y Ciencias Hídricas

Santa Fe, 16 de marzo de 2018.

Como miembros del Jurado Evaluador de la Tesis de Doctorado en Ingeniería titulada **“Reconocimiento de actividades a partir de señales inerciales y acústicas”**, desarrollada por el Bioing Sebastián Rodrigo VANRELL en el marco de la Mención “Inteligencia Computacional, Señales y Sistemas”, certificamos que hemos evaluado la Tesis y recomendamos que sea aceptada como parte de los requisitos para la obtención del título de Doctor en Ingeniería.

La aprobación final de esta disertación estará condicionada a la presentación de dos copias encuadradas de la versión final de la Tesis ante el Comité Académico del Doctorado en Ingeniería.

.....  
Dr. Omar Chiotti

.....  
Dr. Pablo Mandolesi (K)

.....  
Dr. Mariano Rubiolo

.....  
Dr. Humberto Torres

(\*) El Dr. Mandolesi participó por video conferencia

Dr. JOSÉ LUIS MACOR  
SECRETARIO DE POSGRADO  
Facultad de Ingeniería y Cs. Hídricas

Santa Fe, 16 de marzo de 2018

Certifico haber leído la Tesis, preparada bajo mi dirección en el marco de la Mención “Inteligencia Computacional, Señales y Sistemas” y recomiendo que sea aceptada como parte de los requisitos para la obtención del título de Doctor en Ingeniería.

.....  
Dr. Diego Milone  
Codirector de Tesis

.....  
Dr. Leonardo Rufiner  
Director de Tesis

Universidad Nacional del  
Litoral  
Facultad de Ingeniería y  
Ciencias Hídricas

Secretaría de Posgrado

Ciudad Universitaria  
C.C. 217  
Ruta Nacional Nº 168 - Km. 472,4  
(3000) Santa Fe  
Tel: (54) (0342) 4575 229  
Fax: (54) (0342) 4575 224  
E-mail: posgrado@fich.unl.edu.ar





## DECLARACIÓN DEL AUTOR

Esta tesis ha sido remitida como parte de los requisitos para la obtención del grado académico de Doctor en Ingeniería - Mención Inteligencia Computacional, Señales y Sistemas ante la Universidad Nacional del Litoral y ha sido depositada en la Biblioteca de la Facultad de Ingeniería y Ciencias Hídricas para que esté a disposición de sus lectores bajo las condiciones estipuladas por el Reglamento de la mencionada Biblioteca.

Citaciones breves de esta tesis son permitidas sin la necesidad de un permiso especial, en la suposición de que la fuente sea correctamente citada. Solicitudes de permiso para la citación extendida o para la reproducción parcial o total de este manuscrito serían concebidos por el portador legal del derecho de propiedad intelectual de la obra.



*A mi familia*



# Agradecimientos

Muchos e inenerrables.



# Índice general

Resumen . . . . .	XIX
Abstract . . . . .	XXI
1. Introducción . . . . .	1
2. Objetivos . . . . .	4
2.1. Objetivo general . . . . .	4
2.2. Objetivos particulares . . . . .	4
3. Antecedentes . . . . .	5
3.1. Captura de señales, preprocesamiento y análisis por tramos . . . . .	9
3.2. Extracción de características . . . . .	12
3.3. Clasificación y medidas de desempeño . . . . .	13
4. Aportes originales y metodología . . . . .	15
4.1. Señales de entrada, preprocesamiento y análisis por tramos . . . . .	15
4.2. Extracción de características . . . . .	20
4.3. Clasificación y posprocesamiento . . . . .	24
4.4. Validación y medidas de desempeño . . . . .	26
5. Resultados . . . . .	29
5.1. Reconocimiento de actividades humanas . . . . .	30
5.2. Reconocimiento de actividades en vacas lecheras . . . . .	32
6. Conclusiones y trabajos futuros . . . . .	35
Anexos . . . . .	45
A. Assessment of homomorphic analysis for human activity recognition from acceleration signals . . . . .	47
B. 3D acceleration for heat detection in dairy cows . . . . .	73
C. A regularity-based algorithm for identifying grazing and rumination bouts from acoustic signals in grazing cattle . . . . .	91
D. Proceso para detectar celo en un animal rumiante y dispositivo . . . . .	121





# Índice de figuras

1. Ubicaciones típicas de los sensores en el reconocimiento de actividades humanas. . . . .	6
2. Sensores para el reconocimiento de actividades animales. (a) banda de presión (b) micrófono. . . . .	8
3. Esquema general de un reconocedor de actividades. . . . .	9
4. Registro de una (a) secuencia de actividades, (b) el bloque de una actividad y (c) dos acciones (pasos) dentro de la actividad. . . . .	11
5. Fases de análisis de la señal de audio para obtener los bloques de actividad. . . . .	19
6. Obtención de los coeficientes cepstrales. (a) Señal de aceleración cuasi-periódica $a(t)$ ; (b) espectro de $a(t)$ ; (c) logaritmo del espectro; (d) coeficientes cepstrales; (e) aproximaciones de $\log  H(f) $ recuperadas. . . . .	22
7. Sonidos registrados durante a) el pastoreo y b) la rumia. c) Energía de la señal sonora durante el pastoreo (línea de puntos) y la rumia (línea continua). Se indican con flechas rojas las interrupciones relacionadas a la regurgitaciones del bolo alimenticio. . . . .	24
8. Esquema simplificado de la validación cruzada. . . . .	27
9. Posibles situaciones de la comparación entre la secuencia de referencia y reconocida, (a) conservando la relación uno a uno entre bloques o (b) sin relación uno a uno entre bloques. . . . .	28
10. Error en la estimación de los tiempos de rumia para RAFAR-MBBP y el sistema Hi-Tag. El eje superior está normalizado por la longitud de los intervalos analizados (2 horas). . . . .	34



# Índice de tablas

1. Señales inerciales, condiciones de captura y preprocesamiento . . . .	10
2. Características utilizadas en el reconocimiento de actividades. . . .	13
3. Clasificadores seleccionados para el reconocimiento de actividades. .	14
4. Bases de datos disponibles de señales inerciales (A: acelerómetro, G: giróscopo, M: magnetómetro) . . . . .	16
5. Variantes del algoritmo para el reconocimiento de actividades ali- mentarias. . . . .	26
6. Subcategorías de los segmentos identificados inicialmente como VP, FP, VN o FN. . . . .	29
7. Desempeño global comparado mediante el error de reconocimiento.	30
8. Matriz de confusión para el vector de características $F_{24}$ con 70 CCs.	31
9. Comparación de las distintas variantes de la etapa de clasificación. Los valores de las celdas corresponden a la medida-F1, calculada por ventana. . . . .	32
10. Resumen de la segmentación y la clasificación de actividades alimen- tarias. La medida-F basada en ventanas y en bloques se promedió a lo largo de todos los registros analizados (desvío estándar). . . . .	33



# Resumen

El problema de reconocimiento de actividades consiste en identificar las actividades que realiza una persona o animal a lo largo del tiempo. Se presentan diversos desafíos para discriminar entre actividades, como lo son el procesamiento de registros de larga duración y lidiar con información dada en diferentes escalas temporales. Un sistema de reconocimiento automático puede asistir en estas tareas mediante varias modalidades de captura. En particular, en esta tesis se trabaja en el reconocimiento de actividades a partir de señales inerciales y acústicas. Estas se capturan con sensores dispuestos sobre el cuerpo de la persona o el animal, permitiendo el monitoreo continuo y sin restringir el área de captura. A lo largo de las aplicaciones se evalúan diferentes propuestas para el procesamiento de las señales y la clasificación de las actividades.

Las señales inerciales permiten capturar indirectamente los movimientos del cuerpo y se utilizan tanto en humanos como en vacas lecheras. En el caso de las actividades humanas se trabaja principalmente en una nueva representación de las señales, basada en el análisis cepstral. Esta representación permite capturar la dinámica del cuerpo asociada a las actividades de interés. Las características cepstrales, solas o en combinación con otras ya conocidas, se desempeñan muy bien en el reconocimiento de actividades cotidianas y superan a las características del estado del arte. En un esquema similar de trabajo, y de acuerdo a experimentos preliminares, el procesamiento de las señales inerciales también permite la detección de celo en vacas lecheras.

Por otra parte, se utilizan señales acústicas para el monitoreo de las actividades alimentarias en vacas lecheras. Se reconocen las dos actividades principales a partir del procesamiento de las señales sonoras. En una primera etapa, se delimitan los bloques de actividad al detectar la periodicidad de los eventos masticatorios. Esta periodicidad es detectada mediante la autocorrelación de la envolvente de la señal. En una segunda etapa, se clasifican los bloques de actividad caracterizando las interrupciones, las cuales son detectadas a partir de cambios en la energía del sonido. La evaluación de la propuesta se realiza sobre registros de larga duración y utilizando métricas multidimensionales. Los resultados son satisfactorios y superan a un sistema comercial muy utilizado en distintos establecimientos alrededor del mundo.



# Abstract

Activity recognition aims at identifying the activities that a person or animal performs over time. Several challenges are faced in the discrimination of activities, such as processing long-term recordings and dealing with information given at different timescales. An automatic recognition system can assist in these tasks through various modalities. In particular, this thesis deals with activity recognition from inertial and acoustic signals. These signals are captured with sensors on the body of a person or an animal, which allows performing continuous monitoring and do not impose constraints on the capture area. Different proposals for signal processing and activity classification are evaluated along the applications.

Inertial signals indirectly capture body movements and are used in both humans and dairy cows. Regarding human activities, a new representation of signals was studied, based on cepstral analysis. This representation captures the body dynamic that is associated with the activities of interest. Cepstral features, solely or in combination with others already known, perform very well in the recognition of daily activities and outperform state-of-the-art features. In a similar scheme, preliminary experiments showed that processing inertial signals helps to detect the onset of estrus in dairy cows.

On the other hand, acoustic signals are used to monitor foraging behavior in dairy cows. Cattle foraging behavior is mainly composed of grazing and rumination times. An algorithm is proposed for long-term analysis of this behavior to recognize grazing and rumination blocks. In a first stage, a complete recording is analyzed to detect regular masticatory events and to define the time boundaries of foraging activity blocks. The detection of event regularity is based on the autocorrelation of the sound envelope. In a second stage, the energy of acoustic signals within a block is analyzed to detect interruptions and characterize their regularity. Rumination blocks present regular interruptions, whereas grazing blocks do not. An extensive evaluation of the proposed algorithm is made with a set of multidimensional metrics. The results obtained are very good and outperform a commercial system which is widely used in many dairy farms around the world.





# 1. Introducción

Reconocer las actividades desarrolladas por un individuo, o su comportamiento, resulta de interés para una gran variedad de disciplinas. Por ejemplo, en el ámbito de la salud humana para analizar la recuperación de un paciente, en el deporte de alto rendimiento para la asistencia en el entrenamiento, o en el monitoreo de animales para realizar un seguimiento de su bienestar. En ocasiones, una persona podría realizar la tarea de monitoreo mediante observación directa. Sin embargo, en la práctica este enfoque requiere una dedicación exclusiva que la convierte en una tarea costosa y tediosa; y en la mayoría de los casos resulta impracticable. Además, la zona de vigilancia queda restringida al área observable y, cuando se desea realizar el monitoreo simultáneo de varios individuos en un entorno libre, ya no es una tarea posible para una sola persona. Por ello, resulta de interés el desarrollo de sistemas automáticos que permitan reconocer actividades mediante otros enfoques.

En los últimos años se han realizado muchas propuestas para el reconocimiento de actividades, principalmente en el reconocimiento de actividades humanas, aunque también las hay en animales. Dada la diversidad de escenarios y aplicaciones posibles, no dejan de incrementarse los desafíos en el reconocimiento de actividades. Se han obtenido muy buenos resultados en experimentos de laboratorio o escenarios controlados. A medida que nos alejamos de ese ideal aparecen nuevos problemas. Claramente, el escenario de interés es otro, donde el individuo o el animal desarrollen sus actividades libremente. Por lo que la entrada del sistema no debería imponer restricciones sobre el área de captura. Para ello, es posible trabajar con señales capturadas con algún sensor dispuesto sobre el cuerpo del individuo o el animal, que le permitan actuar de forma libre y natural.

Al diseñar y construir un reconocedor, otro de los desafíos es contar con información suficiente acerca de las actividades que se quiere identificar. En general, se cuenta con mayor información de actividades frecuentes. Por ejemplo, en el reconocimiento de las actividades humanas, ha sido ampliamente caracterizada la actividad de caminar. En cambio, hay otras actividades que son eventuales y se cuenta con menos información de ellas. Un ejemplo de estas últimas son las caídas, que son difíciles de registrar en condiciones naturales, no estando la información de

ellas siempre disponible. También es de esperar que durante el monitoreo aparezcan actividades desconocidas, que deben ser distinguidas de las actividades de interés. El sistema tiene que estar preparado para lidiar con los tres tipos de actividades: frecuentes, eventuales y desconocidas. Además se debe considerar la necesidad de ajustar el reconocedor al comportamiento de cada individuo, lo cual puede requerir información extra para construir el reconocedor. Este ajuste al individuo es una estrategia clásica en el reconocimiento automático, y prácticamente en todas las áreas de aplicación se han propuesto sistemas generales o adaptados al individuo.

El sistema de reconocimiento debe ser capaz de identificar actividades con distintas dinámicas. Algunas actividades pueden considerarse estáticas, por no presentar grandes variaciones durante su desarrollo (permanecer de pie o descansando). En cambio otras se componen de una secuencia de acciones o movimientos, que pueden repetirse en múltiples ciclos. Por ejemplo, los pasos que componen el caminar, o las masticaciones en la ingesta de alimento. En ambos casos se presenta una ritmicidad de las acciones que dan origen a la actividad. Otro aspecto a considerar es la duración de las actividades, ya que no todas tienen una duración típica. Por ejemplo, una persona puede caminar por algunos segundos y ser una actividad de corta duración, o puede hacerlo por varios minutos y ser una actividad prolongada. En cambio, sería extraño que una persona suba las escaleras durante un tiempo prolongado. Es así como otros de los desafíos del problema es lidiar con actividades y señales de longitud variable.

La información de las actividades también se da en múltiples escalas temporales. Hay información de corta duración relacionada a acciones individuales, e información de largo plazo, relacionada a las actividades en sí mismas. Aún bajo la consideración de que las acciones sean capturadas perfectamente, reconocer la actividad a partir de ellas no es una tarea trivial. En ocasiones los patrones en la señal exponen claramente una actividad. Otras veces las relaciones que guardan las acciones pueden ser más complejas, involucrando vínculos temporales o una secuencia particular de ellos. Inclusive, en algunos escenarios la relación entre las acciones manifestadas y la actividad subyacente no se conoce completamente. La capacidad de tomar información cruda, reconocer patrones espacio-temporales y tomar una decisión a partir de ello es lo que deseamos que posea el reconocedor. Esta es una tarea compleja que a veces queda eclipsada por la facilidad con la que los humanos reconocemos patrones.

La definición del problema general de reconocimiento de actividades se puede acotar de acuerdo a la naturaleza de los registros de entrada y cuál sea la salida deseada. Con respecto a la entrada, el primer caso corresponde a registros que se encuentren previamente segmentados por actividad. Es decir, donde la señal de cada registro se corresponde a una única actividad. El segundo caso es aquel donde cada registro contiene una secuencia de actividades. Esta secuencia puede estar

integrada exclusivamente por actividades de interés o también incluir momentos de inactividad u otra actividad irrelevante. En todos los casos la señal dentro de cada registro es continua y la duración puede variar de registro a registro, sin que ello modifique el tipo de problema. En general, los reconocedores procesan y clasifican la señal por tramos, obteniendo una secuencia de tramos etiquetados que se corresponde con el registro de entrada. La salida deseada puede ser una secuencia de etiquetas que abarque la duración completa de cada registro o simplemente una etiqueta de actividad para el registro completo.

La combinación de entrada y salida permite vincular el reconocimiento de actividades a tipos de problemas conocidos dentro del reconocimiento automático a partir de señales temporales. Este vínculo permite guiar no sólo el desarrollo de un reconocedor, sino también la evaluación del mismo. Un problema de clasificación clásico es aquel donde la salida es única (una etiqueta de actividad). Para evaluar un reconocedor en este problema se dispone de una gran variedad de medidas. Por ejemplo, la tasa de acierto, la precisión y la sensibilidad. En cambio, en el caso en que las salidas sean secuencias de etiquetas, la evaluación del reconocedor no es tan simple, dado que las alternativas propuestas en otras áreas para problemas similares no se ajustan completamente a este problema.

En esta tesis se trabaja partiendo desde un enfoque general, para luego aplicarlo al reconocimiento de actividades en problemas específicos. Se trabaja en el reconocimiento de actividades a partir de sensores inerciales o acústicos dispuestos sobre el cuerpo del individuo o el animal. Los sensores inerciales brindan información indirecta de los movimientos y son utilizados tanto en humanos como en vacas lecheras para el reconocimiento de actividades. Los sensores acústicos son utilizados para monitorear las actividades alimentarias en vacas lecheras. El trabajo se centra sobre las técnicas de procesamiento de señales para extraer información relevante de las mismas, y sobre los algoritmos heurísticos y de aprendizaje maquina que puedan explotar dicha información para reconocer las actividades de interés. En cada aplicación se busca desarrollar un sistema que no necesite ser adaptado al individuo o animal para desempeñarse adecuadamente. Se proponen y evalúan alternativas a las técnicas conocidas, en la búsqueda de mejorar el desempeño en los escenarios que hayan sido explorados anteriormente y resolver los nuevos desafíos que se presenten. Además se estudia el problema de evaluar el reconocimiento continuo de actividades, el cual presenta características que le son propias y no se adecúa a los problemas de clasificación estándares. El objetivo final es obtener reconocedores que demuestren ser confiables sobre registros continuos y de larga duración.

## Organización del documento

La presente tesis está organizada bajo el formato de *tesis por compilación*. En la Sección 2 se detallan los objetivos y en la Sección 3 se realiza una revisión de los antecedentes en el reconocimiento de actividades humanas y animales a partir de las señales mencionadas. En la Sección 4, se describen el reconocedor, los métodos propuestos y la metodología general aplicada. La Sección 5 describe brevemente las aplicaciones y los resultados. Las conclusiones generales junto con los trabajos futuros se dan en la Sección 6. Finalmente, los artículos en los anexos describen en detalle los problemas particulares, las propuestas, los experimentos y los resultados obtenidos. El Anexo A se corresponde a la propuesta de las características cepstrales, los Anexos B y D al detector de celo, y el Anexo C al reconocedor de actividades alimentarias a partir de señales sonoras.

## 2. Objetivos

### 2.1. Objetivo general

El objetivo general de esta tesis es proveer nuevos algoritmos que sean capaces de reconocer actividades a partir de señales capturadas con sensores inerciales y acústicos dispuestos sobre el cuerpo del individuo o animal.

### 2.2. Objetivos particulares

Los objetivos particulares están relacionados con las etapas necesarias para obtener un reconocedor automático de actividades:

1. Analizar las bases de datos disponibles que cuenten con señales inerciales producidas por los movimientos de un individuo o animal realizando diversas actividades, para seleccionar las más adecuadas.
2. Analizar las bases de datos disponibles que cuenten con señales acústicas correspondientes al comportamiento alimentario animal, para seleccionar la más adecuada.
3. Proponer y evaluar nuevas técnicas para el procesamiento y análisis de señales capturadas con sensores inerciales o acústicos con el fin de extraer características que brinden información relevante y discriminativa sobre las actividades desarrolladas.
4. Evaluar clasificadores que permitan reconocer las actividades desarrolladas por un individuo o animal a partir de las características extraídas.

5. Analizar el desempeño de los algoritmos propuestos mediante la utilización de tests estadísticos y métodos de validación.
6. Analizar la extensión de las técnicas propuestas al reconocimiento continuo de actividades en señales de larga duración.
7. Aplicar las técnicas desarrolladas para el reconocimiento de actividades humanas y de comportamientos animales, de interés en el área de computación ubicua y en la ganadería de precisión, respectivamente.

### 3. Antecedentes

El reconocimiento de actividades ha sido objeto de estudio en diversas disciplinas y es un tema de creciente interés científico [1]-[7]. Mayormente se ha estudiado el reconocimiento de actividades humanas y en menor medida el reconocimiento de actividades animales. En cuanto a esta última, son de nuestro interés las aplicaciones relacionadas a la ganadería de precisión. Todos estos trabajos muestran que principalmente existen dos estrategias para capturar información sobre las acciones y actividades que realiza una persona o un animal. La primera consiste en registrar las acciones o movimientos con un dispositivo o sensor externo al cuerpo. La segunda estrategia emplea uno o más dispositivos sobre el cuerpo de la persona o el animal para realizar la captura.

Entre los sensores externos, la captura en video es la más popular, aunque también se han utilizado cámaras de profundidad, solas o en combinación con videos [1], [2], [7], [8]. Para la captura precisa de los movimientos también se han utilizado múltiples cámaras fijas y marcadores sobre el cuerpo, aunque no resultan prácticos para el reconocimiento de actividades en general [9].

En el caso de los dispositivos ubicados sobre el cuerpo, se han utilizado mayormente sensores inerciales [3]-[5], [10]-[16]. Éstos incluyen acelerómetros, giroscopios y brújulas, siendo los dos primeros los más útiles para el reconocimiento. En ocasiones, la información inercial se ha combinado con el registro simultáneo de la geolocalización del individuo [14], [17]. Sin embargo, en los trabajos más recientes se observa una tendencia a no utilizar la geolocalización como fuente de información primaria para el reconocimiento, dado que la dinámica de las actividades suele superar la resolución temporal de la geolocalización. Otros sensores que han sido utilizados como complemento a los inerciales, aunque en menor medida, son los micrófonos, sensores de presión, sensores de ritmo cardíaco y galgas extensiométricas [3], [7], [12], [18]. La información multimodal ha demostrado mejorar los resultados obtenidos exclusivamente con sensores inerciales. Sin embargo, esta mejora está acompañada de un lógico aumento en la complejidad para la captura y procesamiento de las señales, lo que no siempre se justifica a nivel práctico. En

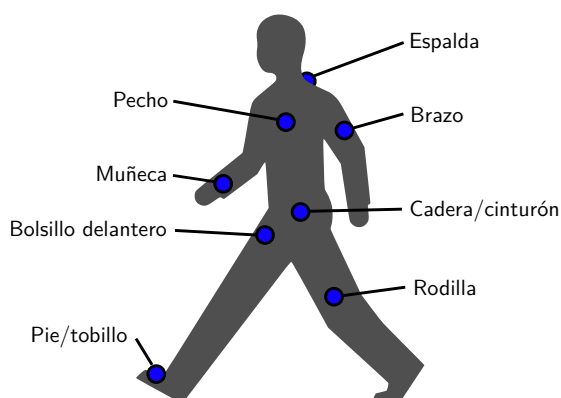


Figura 1: Ubicaciones típicas de los sensores en el reconocimiento de actividades humanas.

cambio, los mismos sensores en solitario han sido utilizados con otros fines como lo son el reconocimiento de emociones o estados de alerta, reconocimiento automático del habla, monitoreo de la respiración o estimación del esfuerzo físico [12].

Además de que distintos tipos de sensores pueden concentrarse en un mismo dispositivo, en otros estudios se analizó la captura simultánea con más de un dispositivo sobre el cuerpo del individuo [19]-[21]. En la Figura 1 se muestran algunas ubicaciones que han sido estudiadas en trabajos previos sobre el reconocimiento de actividades humanas. La cintura, el pecho, los brazos y los antebrazos son de las posiciones más habituales. En las aplicaciones ganaderas también se han considerado diversas ubicaciones, siendo los lugares habituales el cuello, la cabeza y la pata del animal [7], [10], [14]. Las propuestas con dispositivos en más de una ubicación simultánea suelen contar con un dispositivo central, el cual recibe información de los restantes ubicados en las extremidades. Por supuesto, a mayor número de dispositivos la complejidad para la captura y procesamiento aumentan considerablemente. Dependiendo de las actividades de interés, la utilización de más de un dispositivo podría resultar en un marcado beneficio. Por ejemplo, en [21] se estudió el reconocimiento de actividades complejas como preparar el desayuno, donde las actividades consisten en una persona que interactúa con ambas manos y distintos objetos.

Cada estrategia de captura, sobre el cuerpo o externa a él, tiene sus ventajas y desventajas que son más o menos importantes de acuerdo al campo de aplicación y al objetivo buscado. Por ejemplo, la captura con cámaras de video tiene la ventaja de no requerir dispositivo alguno sobre el cuerpo de la persona o el animal. Esta estrategia ha sido utilizada para detectar signos del celo en vacas en rodeos confinados [8]. Sin embargo, para el reconocimiento de actividades en general, los sensores dispuestos sobre el cuerpo poseen varias ventajas respecto a la captura

mediante cámaras de video [3], [4], [7], [11], [13]. Entre éstas se puede mencionar que los sensores sobre el cuerpo permiten que la captura se realice tanto en ambientes cerrados como abiertos, sin restricciones a que el individuo o animal permanezca en una zona o región particular. En el caso de la detección de signos de celo, permitiría realizar el monitoreo en rodeos a campo abierto. Además, los sensores sobre el cuerpo son de reducido tamaño, consumen poca energía, y pueden ser incorporados en un dispositivo portátil, de fácil sujeción al cuerpo y mínima invasividad. De esta manera, la persona o el animal bajo análisis puede desarrollar sus actividades sin verse incomodado por el dispositivo.

La utilización de los acelerómetros para el reconocimiento de actividades humanas se inició en la década de los 90 [22], [23]. En aquellos estudios se logró la distinción entre actividades estáticas y dinámicas. Hoy en día, los reconocedores de actividades utilizan acelerómetros u otros sensores triaxiales y deben reconocer un conjunto más amplio de actividades [3], [12], [20], [24]-[26]. En el caso del reconocimiento de actividades animales, también se han utilizado acelerómetros con cierto éxito para identificar actividades locomotoras [10], [14], [27] y no tanto para las alimentarias [7]. Sin embargo, la variedad de técnicas exploradas para el reconocimiento de actividades humanas supera ampliamente a las que se han aplicado para el reconocimiento de actividades animales.

En cuanto al monitoreo de las actividades alimentarias en rumiantes, los sensores inerciales han sido superados por el uso de otros sensores como bandas de presión y micrófonos [7], [28]-[31]. Uno de los problemas con los sensores inerciales es que por medio del análisis de la postura del animal se busca reconocer tanto actividades alimentarias como de locomoción. Sin embargo, hay actividades que comparten posturas similares y se confunden entre sí. Por ejemplo, los rumiantes mientras permanecen de pie o recostados pueden estar reprocesando los alimentos previamente ingeridos. Por lo tanto, detectar que el cuerpo se encuentra inmóvil no es suficiente para discriminar las actividades posibles en esa condición.

Las bandas de presión alrededor del hocico (Figura 2.a) resultan en una mejor estrategia para el reconocimiento de actividades alimentarias, dado que permiten reconocer exclusivamente movimientos de la mandíbula [28], [31]. Estas pueden capturar los movimientos y así reconocer las actividades de ingestión del alimento, la rumia y los momentos en que el animal bebe. Sin embargo, los estudios con estas técnicas se limitaron al análisis de un máximo de dos horas continuas, aún cuando las actividades alimentarias en rumiantes pueden ser más extensas y ocupar la mayor parte del día [32].

Mediante el uso de micrófonos (Figura 2.b) se pueden capturar señales de audio con el fin de caracterizar y cuantificar el comportamiento alimentario en vacas [29], [30], [33], [34] y ovejas [35], [36]. Estas técnicas han demostrado ser muy buenas para el monitoreo de los eventos en el corto plazo, especialmente en campo abierto.

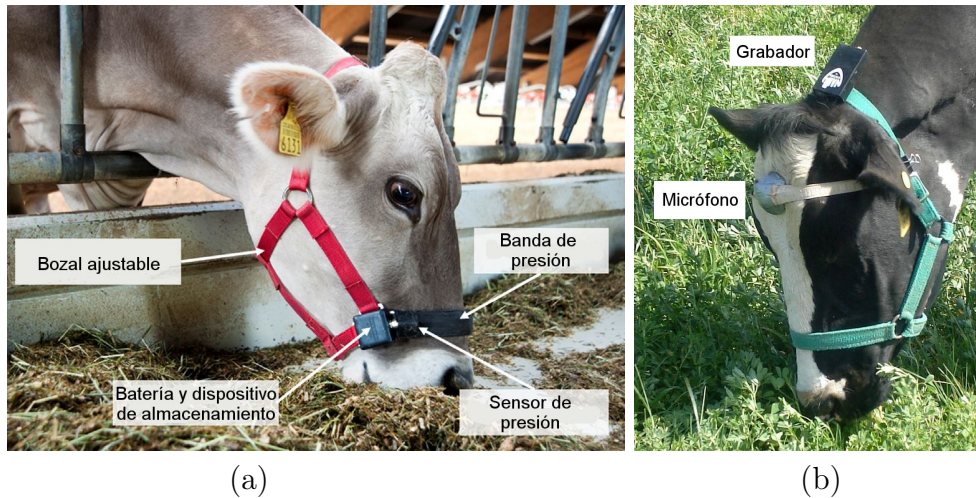


Figura 2: Sensores para el reconocimiento de actividades animales. (a) banda de presión (b) micrófono.

Sin embargo, aún no se ha estudiado su extensión para el monitoreo continuo y a largo plazo de las actividades alimentarias. Existe un sistema comercial de monitoreo, Hi-Tag [37], [38], que se basa en un micrófono interno. Sin embargo, este sólo se enfoca en reconocer la rumia y ha sido evaluado sólo en rodeos confinados, por lo que no se conoce su funcionamiento en condiciones de campo abierto.

Como la mayoría de los sistemas automáticos de reconocimiento, los reconocedores de actividades pueden considerarse compuestos por bloques como los mostrados en la Figura 3. En la captura un sensor convierte un desplazamiento, una aceleración, un sonido, u otra variable física en la señal de entrada. Es posible que la señal de entrada contenga ruido o que limite las posibilidades del procesamiento en las etapas siguientes. En este caso, como bloque inicial del reconocedor se ubica una etapa de preprocesamiento que busca acondicionar la señal de entrada. Luego, se parte la señal para su análisis por tramos y, de ser posible, se identifican y descartan los períodos que no corresponden a ninguna actividad de interés. En la extracción de características se obtienen propiedades de los tramos de la señal que permitan realizar la clasificación. El clasificador, a partir de estas características, utiliza sus modelos de actividad para definir cual le corresponde al tramo de señal. Finalmente, el posprocesamiento puede considerar información adicional del entorno o del dominio de aplicación para modificar la salida del clasificador y brindar la salida final. Esta salida será una etiqueta de actividad en el caso de los registros segmentados previamente o una secuencia de actividades en el caso de los registros continuos. Si bien esta descripción define un solo sentido de la información (desde la señal de entrada hacia la actividad final), un reconocedor



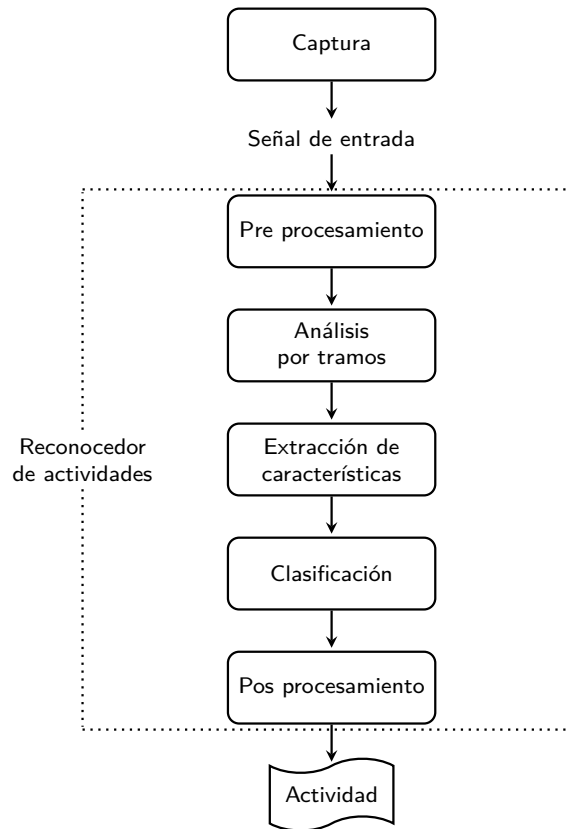


Figura 3: Esquema general de un reconocedor de actividades.

podría emplear alguna realimentación de niveles inferiores hacia los superiores, o intercalar etapas de procesamiento entre los bloques mencionados. Para mejorar cada uno de estos bloques se han realizado muchas propuestas en las últimas décadas [3], [4], [7], [12], [19], [39]-[41]. A continuación se resumen las más relevantes en relación a los objetivos planteados.

### 3.1. Captura de señales, preprocesamiento y análisis por tramos

Los trabajos en el reconocimiento de actividades han estudiado diferentes configuraciones para la captura de señales inerciales. En la Tabla 1 se detallan las configuraciones de algunos trabajos de referencia, para dar cuenta de esta diversidad. En cuanto a los sensores utilizados, la mayoría de los trabajos han optado por los acelerómetros. En ocasiones, incluso en dos o más ubicaciones [20], [23], [26]. Los giróscopos han sido incorporados en trabajos más recientes [42], [43]. Su incorporación trae beneficios en el reconocimiento a costa de un aumento en el

Tabla 1: Señales inerciales, condiciones de captura y preprocesamiento

Acelerómetro	Giroscopio	Frecuencia muestreo	Ancho ventana	Filtro pasa bajos	Escenario	Trabajo	
2*	1D	10 Hz	10.00 s	0.5 Hz	Lab	[23]	
5*	2D	76 Hz	6.70 s	No	Semi Nat	[20]	
1*	3D	10 Hz	10.00 s	No	Semi Nat	[10]	
3*	3D	100 Hz	5.12 s	No	Nat	[26]	
1*	3D	1* 3D	50 Hz	2.56 s	0.3 Hz	Lab	[42]
1*	3D		10 Hz	10.00 s	No	Nat	[45]
1*	3D	1* 3D	50 Hz	2.56 s	0.3 Hz	Lab	[43]
1*	3D	1* 3D	100 Hz	4.00 s	No	Lab+Nat	[24]
1*	3D	1* 3D	100 Hz	1-10 s	1-2 Hz	Nat	[44]

costo de procesamiento. Con respecto a la frecuencia de muestreo, se han elegido mayormente entre 50 y 100 Hz, superando los 20 Hz mínimos necesarios para la captura sin distorsión de actividades humanas cotidianas [20]. En cambio, para la detección de actividades animales se han utilizado frecuencias de muestreo de entre 10 y 100 Hz [10], [44], [45]. Los escenarios de captura han ido evolucionando desde los experimentos muy controlados o montados en laboratorios (Lab) a otros más naturales (Nat). Uno de los desafíos actuales es extender a condiciones naturales, los buenos resultados obtenidos en condiciones de laboratorio.

El bloque de preprocesamiento busca eliminar artefactos que pudieran incorporarse a la señal antes de extraer información de ella. Por ejemplo, cuando se desea analizar la aceleración del cuerpo del individuo, la aceleración gravitacional es un artefacto que se mezcla en la señal registrada. Si bien varios trabajos han hecho esfuerzos por eliminarla [10], [42], [43], esto no se logra completamente y el sistema debe ser robusto a su presencia. En la Tabla 1 se indican las frecuencias de corte de los filtros pasa bajos (PB) que han utilizado algunos trabajos con este fin [23], [42]-[44]. Los beneficios de aplicar este proceso no están claramente establecidos.

En el caso de las señales de audio para el monitoreo del comportamiento alimentario, la captura se realizó en las frecuencias habituales, entre 8 y 44 kHz. Por ello, la etapa de preprocesamiento suele involucrar una etapa de filtrado pasa bajos o pasa banda que sólo conserve información sobre los eventos que interesa detectar pero no así sobre otros sonidos o ruidos ambientes que pudieran quedar registrados [30], [34], [36], [46]. En el caso de la detección y clasificación de eventos masticatorios, los cuales se suceden a una frecuencia cercana a 1 Hz, el preprocesamiento involucra también la obtención de la envolvente de la señal de audio, muestreada a una frecuencia varios órdenes de magnitud por debajo de la original [33]. Esto permite reducir enormemente el costo computacional de las etapas siguientes del reconocedor.

La tarea que le sigue al preprocesamiento es la división de los registros origina-

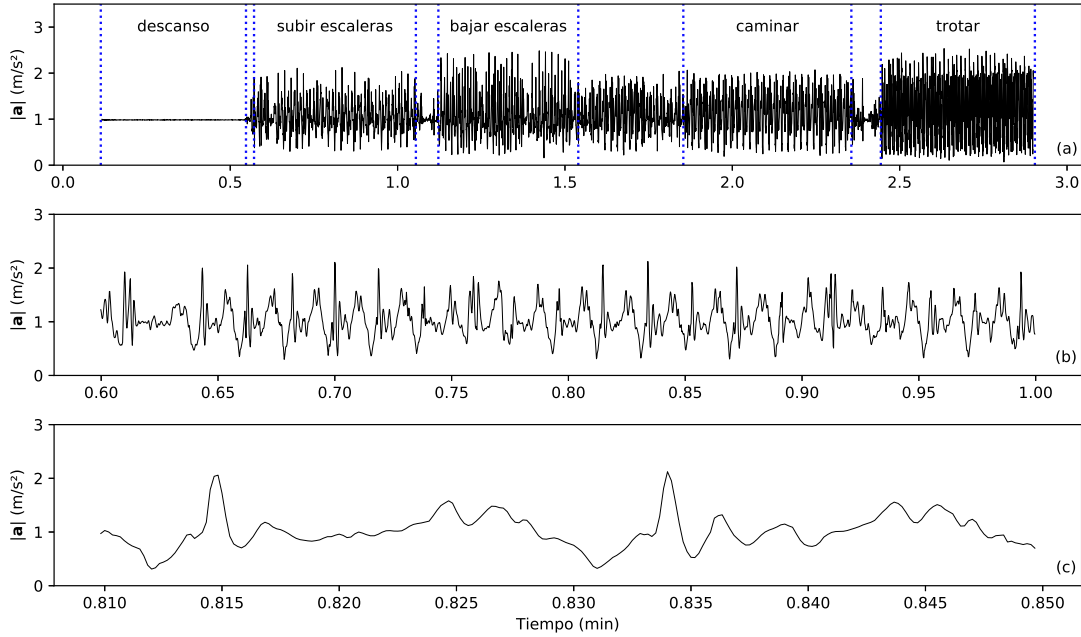


Figura 4: Registro de una (a) secuencia de actividades, (b) el bloque de una actividad y (c) dos acciones (pasos) dentro de la actividad.

les para su posterior análisis [12], [19], [41], [47]. La estrategia consiste en dividir la señal en tramos a partir de establecer bloques o ventanas de análisis. Estos bloques pueden estar predefinidos o ser generados a partir de la detección de eventos o de cambios en la señal. La mayoría de los reconocedores trabajan mediante una ventana deslizante de tamaño fijo con algún grado de solapamiento, sin importar como se compongan las etapas siguientes [10], [20], [24], [26]. El tamaño de la ventana suele estar en el orden de los segundos (Tabla 1), acorde con la duración de las acciones características de las actividades que se desean reconocer. En aplicaciones de tiempo real, las ventanas permiten procesar continuamente la señal, a medida que se va capturando, y generar una salida [33]. Cuando el procesamiento no se realiza en tiempo real, puede dividirse el registro completo y luego analizar las ventanas secuencialmente. En el caso de la división en tramos basada en la detección de eventos puntuales, los mismos se utilizan para definir los límites de las ventanas o los bloques. Esta estrategia da como resultado ventanas o bloques que no necesariamente tendrán un tamaño fijo. Finalmente, mediante la detección de cambios de patrón en la señal se puede identificar los momentos en que ocurre una transición de un patrón a otro. Por ejemplo, en la Figura 4 se observan los cambios de patrón en la señal de aceleración a lo largo de una secuencia de actividades. Estos momentos de cambios se pueden utilizar para definir los límites de los

bloques [12], [19]. Algunas alternativas para la detección de cambios son: analizar la energía, la entropía o complejidad de la señal, utilizar la señal de una modalidad para segmentar otra o, inclusive, utilizar una fuente de información externa [48].

### 3.2. Extracción de características

En trabajos previos se ha utilizado una gran variedad de técnicas para la extracción de características a partir de los sensores mencionados (Tabla 2). Agrupando las principales técnicas de acuerdo al tipo de características resultantes nos encontramos con: las heurísticas, las temporales, las frecuenciales y las que están en el dominio tiempo-frecuencia [19]. Otras taxonomías de las características extraídas en este área pueden encontrarse en [3] y [12]. Las características heurísticas se basan en un entendimiento intuitivo de los efectos que una actividad provoca en la señal. Por ejemplo, una actividad estática, como estar de pie o acostado, puede reconocerse por la dirección en la que se registra la gravedad (única aceleración presente) [49]. Esta misma señal puede utilizarse para reconocer la posición relativa o inclinación de la cabeza de un animal, y así estimar el momento en que está pastando [31], [44]. Otros ejemplos de características heurísticas son el área bajo la señal de magnitud, la aceleración pico a pico y la raíz cuadrática media, medidas que en todos los casos han sido relacionadas a la intensidad de una actividad [50]. En el monitoreo del comportamiento alimentario en rumiantes los eventos masticatorios se han detectado y caracterizado mediante medidas heurísticas como la duración, la amplitud y la forma de la envolvente [33].

A diferencia de las heurísticas, las características temporales no tienen una relación directa con una acción o evento específico. En cambio, éstas suelen ser medidas estadísticas de la señal ventaneada. Ejemplos comunes son la media, la varianza, la kurtosis y la correlación entre ejes del sensor [19], [20], [43], [51], [52]. Tanto las características temporales como las heurísticas son las más simples de obtener, y han sido útiles para distinguir entre actividades estáticas y dinámicas. Sin embargo, se cuestiona su efectividad para discriminar entre actividades dinámicas [19], [26]. Su bajo costo computacional las hace atractivas para aquellas aplicaciones donde se pretende implementar el procesamiento de las señales en el mismo dispositivo de captura.

Las características en el dominio tiempo-frecuencia han sido efectivas para detectar las transiciones entre actividades pero no han demostrado ser útiles para clasificar actividades [56], [57]. En cambio, las características frecuenciales son las más utilizadas en los reconocedores de actividades [20], [24], [26], [42], [54], [55]. Estas características se obtienen utilizando la transformada rápida de Fourier (FFT) o la transformada coseno discreta. Los coeficientes de las transformadas se pueden adoptar directamente como características o se puede utilizar algún método adicional, como un banco de filtros o subconjuntos de los coeficientes, para

Tabla 2: Características utilizadas en el reconocimiento de actividades.

Tipo	Ejemplos	Utilidad	Trabajos
Heurísticas	Dirección de la gravedad, área bajo la señal de magnitud, aceleración pico a pico, raíz cuadrática media	Relacionadas a posturas o movimientos específicos; distinción entre actividades estáticas y dinámicas	[11], [23], [31], [44], [49], [50]
Temporales	Media, varianza, kurtosis, energía, correlación entre ejes	Genéricas y de bajo costo computacional	[4], [19], [20], [24], [26], [42], [53]
Frecuenciales	Transformada de Fourier, Entropía del espectro, Transformada Coseno Discreta	Distinción entre actividades dinámicas	[20], [24], [26], [42], [53]-[55]
Tiempo-frecuencia	Transformada wavelet	Detección de transiciones	[56], [57]

la descripción del espectro. Muchos trabajos han reportado sus mejores resultados utilizando como características a los coeficientes de la FFT, individualmente [20], [26], [54], [55] o combinados con características temporales [3], [24], [42], [43]. Sin embargo, no se ha considerado aún cómo incorporar información propia del dominio en esta representación, con el fin de obtener características discriminativas que a la vez puedan interpretarse en términos de las actividades desarrolladas. Recientemente se han propuesto otras técnicas, que consideran características aprendidas automáticamente de los datos. Dichos reconocedores se basan en redes de aprendizaje profundo, como las redes convolucionales [58]-[60]. Sin embargo, estos trabajos coinciden en que se requiere una mayor cantidad de datos para poder evaluar correctamente estas propuestas de arquitecturas complejas, que poseen mayor cantidad de parámetros.

### 3.3. Clasificación y medidas de desempeño

El abanico de técnicas que han sido utilizadas para la etapa de clasificación también es amplio [3], [4], [7], [19]. El empleo de reglas y umbrales ha sido particularmente útil para detectar caídas en humanos a partir de las señales de aceleración pero no así para clasificar actividades [19], [50]. Para clasificar actividades alimentarias en vacas también se han propuesto reglas y umbrales [15], [28], [44]. Para clasificar actividades locomotoras han demostrado mayor éxito las técnicas clásicas de aprendizaje maquina, como árboles de decisión, k vecinos más cercanos, perceptrón multicapa, máquinas de soporte vectorial y bosques aleatorios (Tabla 3). Con cualquiera de ellas la clasificación se realiza por ventanas, es decir, que el clasificador asigna una etiqueta de actividad a cada ventana. En el caso de los registros

Tabla 3: Clasificadores seleccionados para el reconocimiento de actividades.

Clasificador	Trabajos
Árbol de decisión	Bao y col. [20], Zehner y col. [28], Andriamandroso y col. [44], Lara y col. [61] y Parkka y col. [62]
k-vecinos más cercanos	Preece y col. [19] y Reiss y col. [53]
Máquina de soporte vectorial	Martiskainen y col. [10], Xue y col. [26] y Anguita y col. [42]
Random forest	Alvarenga y col. [11] y Ichino y col. [24]

segmentados, en los que todas las ventanas corresponden a una sola actividad, se ponderan las asignaciones por ventana para definir la actividad más probable [3], [19]. La incorrecta clasificación en alguna de las ventanas no es crítica porque ya se sabe que el registro corresponde a una sola actividad. En cambio, cuando el registro involucra una secuencia de actividades, lo que se busca obtener es una secuencia continua de etiquetas de actividades. En este caso, la correcta clasificación en cada ventana es crucial para poder modelar y construir una secuencia lógica de actividades, sin importar el orden en el que se presenten y con el menor número de falsas asignaciones.

Para el modelado de secuencias de actividades también se han propuesto técnicas basadas en modelos ocultos de Markov [43], [63], [64]. Al contrario de las técnicas clásicas mencionadas, estos modelos contemplan en su estructura las probabilidades de transición entre actividades. Algunas de estas propuestas modelan la secuencia de actividades a partir de la secuencia de características extraídas por ventanas [43], [63], mientras que en otros trabajos modelan la secuencia de actividades en un segundo nivel, a partir de la clasificación realizada con algoritmos clásicos [64]. Recientemente también se han propuesto alternativas basadas en redes recurrentes LSTM (del inglés, *long short term memory*) [65], [66]. Éstas buscan incorporar en el aprendizaje de la red las variaciones a corto y largo plazo.

Para evaluar el desempeño en la clasificación de actividades se han utilizado medidas clásicas como la tasa de acierto, sensibilidad, precisión, medida-F, especificidad, área bajo la curva o coeficientes de correlación [10], [19], [24], [26], [28], [44]. En términos generales no existe una medida que sea mejor que otra, cada una contempla los errores de diferente manera, por lo que se recomienda elegir una apropiada combinación de ellas [67]. Algunas pueden aplicarse a los problemas de clasificación binaria o multi-clase, mientras que otras requieren de promedios balanceados para evitar sesgos en los resultados. Es posible encontrar estudios que promedian las medidas por ventanas, por actividades, por sesiones o por sujetos, no habiendo un consenso en cuales medidas seleccionar. Esta diversidad, en parte propia y en parte compartida con otras áreas, impide la comparación justa entre

estudios, que de por sí es compleja, dado que no poseen condiciones experimentales equivalentes. En ocasiones se han propuesto medidas específicas para la segmentación, como la tasa de error de segmentación por actividad [43]. Se han propuesto medidas más elaboradas que proveen diferentes ópticas al clasificar secuencias de actividades [68], [69]. Sin embargo su uso no se ha extendido en el área.

## 4. Aportes originales y metodología

En esta tesis el reconocimiento automático de actividades se llevó adelante mediante un sistema estructurado en bloques, como es habitual en el reconocimiento automático de patrones (Figura 3). A continuación se brinda una descripción general de cada uno de los bloques constituyentes y se detallan las propuestas y adaptaciones particulares para cada aplicación en el reconocimiento de actividades humanas y en vacas lecheras. Finalmente se brinda una descripción de las medidas de desempeño y metodología de validación.

### 4.1. Señales de entrada, preprocesamiento y análisis por tramos

En esta tesis, se propone utilizar dos tipos de señales de entrada, inerciales para la captura de los movimientos en humanos y animales, y acústicas, para la captura de los sonidos masticatorios en animales. En ambos casos las señales fueron capturadas con sensores dispuestos sobre el cuerpo del individuo o el animal. Sin embargo, la captura de las señales no constituye parte de la investigación en sí misma, por lo que se utilizaron señales ya registradas. En el caso de las señales inerciales capturadas en humanos, se analizaron y seleccionaron entre las bases de datos de libre acceso (Tabla 4). En el caso de las señales capturadas sobre animales, se utilizaron bases de datos capturadas por un grupo de la Facultad de Ciencias Agrarias de la Universidad Nacional de Rosario. Hace ya varios años que se trabaja en colaboración con los investigadores de dicho grupo, en temas relacionados a la ganadería de precisión.

#### Reconocimiento de actividades humanas

Para la selección de las bases de datos de señales inerciales se consideraron criterios experimentales y técnicos. Entre los experimentales, se favorecieron aquellas bases de datos que se habían registrado sobre un mayor número de individuos, desarrollando actividades en condiciones naturales y donde las señales estuvieran registradas a una frecuencia suficiente para capturar sin distorsiones la dinámica de los movimientos. Entre los criterios técnicos, el principal fue el de los sensores

Tabla 4: Bases de datos disponibles de señales inerciales (A: acelerómetro, G: giróscopo, M: magnetómetro)

Nombre	sujetos	equipos	actividades	Sensores	Comentarios
UC Berkeley WARD [70]	20	5	12	A+G	señales con datos faltantes
Opportunity [71]	12	5	6	A+G+M	registros continuos, captura en laboratorio
USC- HAD [25]	14	1	9	A+G	registros segmentados
SCUT- NAA [26]	44	3	10	A	registros segmentados
SHAR [72]	30	1	6	A+G	registros continuos, captura en laboratorio
HASC [24]	510	1	6	A+G+M	registros segmentados y continuos

a considerar, siendo indispensable el uso de un acelerómetro y favoreciendo el uso de un sólo dispositivo para la captura de las señales.

En la búsqueda y selección de una base de datos con señales inerciales se encontró que existían varias alternativas de libre acceso vinculadas al reconocimiento de actividades humanas (Tabla 4). Varias de estas bases de datos fueron descartadas de los experimentos por no cumplir con los criterios mencionados. La base de datos UC Berkeley WARD [70] fue obtenida a partir de un diseño experimental adecuado pero muchos de los registros que la integran poseen datos faltantes que, lamentablemente, no permiten su aprovechamiento. Por ejemplo, los datos faltantes impiden que la validación cruzada de los resultados pueda hacerse con particiones equivalentes. De igual manera, las bases de datos Opportunity [71], USC-HAD [25] y SHAR [72] fueron descartadas por haber sido registradas en condiciones de laboratorio (poco naturales) y poseer pocos sujetos. Para el reconocimiento de actividades en señales segmentadas se eligió la base de datos SCUT-NAA [26]. La base de datos HASC [24] se seleccionó para los experimentos con secuencias de actividades continuas. Ambas bases de datos seleccionadas permiten comparar las técnicas propuestas con trabajos recientes que utilizaron los mismos datos.

La base de datos SCUT-NAA [26] cuenta con 1278 registros capturados con acelerómetros, sobre 44 sujetos 30 hombres y 14 mujeres, realizando 10 actividades y a una frecuencia de muestreo de 100 Hz (Tabla 4). Los 44 sujetos tenían entre 21 y 30 años y realizaron cada actividad por al menos 1 minuto. La cantidad de sujetos asegura cierta variabilidad entre individuos (no es el caso en otras bases de datos) y la cantidad de actividades presenta un desafío mayor (en otros casos sólo se consideran 6 actividades). La variedad de actividades provee de ejemplos de baja, media y alta intensidad. Además, la frecuencia de muestreo está muy



por encima de los 20 Hz mínimos necesarios para la captura sin distorsión de actividades cotidianas [20]. El dispositivo de captura estuvo ubicado en la cintura del sujeto durante los registros, la cual es una posición que representa en gran medida los movimientos globales del cuerpo por ubicarse cerca de su centro de masa.

Para el reconocimiento de secuencias de actividades, se buscó que éstas fueran representativas de una situación cotidiana. Además, se buscó que cada registro corresponda a señales capturadas de forma continua, sin secuencias artificiales ni alteraciones de los segmentos de transición, que podrían contener movimientos que el reconocedor debería aprender a ignorar. La base de datos seleccionada HASC [24] cumplió con estos requisitos. Fue creada fruto del trabajo colaborativo de grupos independientes a lo largo de 6 años. Esto ha logrado no sólo que la cantidad de sujetos y sesiones de captura sea mayor sino también que la variabilidad que presenta la base de datos sea representativa de aquellas con la que podría encontrarse el sistema funcionando en el mundo real. Los registros se corresponden a 6 actividades: caminar, permanecer de pie, trotar, trotar dando saltos, subir y bajar escalera. Además, los registros poseen tramos sin etiquetar, que fueron considerados como aquellos que no corresponden a ninguna de las actividades de interés (clase nula).

Para la construcción del reconocedor, las señales de entrada fueron las componentes de la aceleración  $a_x(t)$ ,  $a_y(t)$  y  $a_z(t)$ , en conjunto con la magnitud de la aceleración  $|\mathbf{a}(t)|$ . Para los registros segmentados por actividad (SCUT-NAA), la etapa de preprocesamiento sólo se ocupó de escalar la señal de entrada a una magnitud física estándar. Para el problema de secuencias de actividades (HASC), se decidió considerar la magnitud de la aceleración  $|\mathbf{a}(t)|$  como única señal de entrada, dado los buenos resultados obtenidos en los registros segmentados. Esta señal tiene la ventaja de que es independiente de la orientación del dispositivo de captura. Dado el trabajo colectivo en la formación de esta base de datos, hay señales que fueron registradas a diferentes frecuencias. Por ello, como parte del preprocesamiento también se uniformó el muestreo interpolando a 100 Hz.

Las señales registradas no poseen una longitud fija por lo que, para el análisis por tramos, se eligió utilizar una ventana de 5.12 s, con un solapamiento del 50%. Esto permitió capturar más de un ciclo de las actividades a reconocer. Todas las señales, componentes y magnitud, fueron divididas en ventanas y procesadas de igual manera, por lo que en la descripción de la extracción de características se menciona el procesamiento realizado para una señal ventaneada genérica  $a(t)$ .

## Reconocimiento de actividades en vacas lecheras

Para el reconocimiento de actividades en vacas lecheras se consideraron dos bases de datos, una con señales inerciales y la otra con señales acústicas. La base

de datos de señales inerciales se obtuvo a partir de experimentos con animales en el Campo Experimental de la Facultad de Ciencias Agrarias, Universidad Nacional de Rosario. La base de datos de señales acústicas se registró en la Estación Biológica Kellogg de la Universidad Estatal de Michigan, Estados Unidos. Los experimentos con animales fueron realizados por expertos en el área de producción animal. Para ambos casos los animales involucrados en los experimentos fueron de la raza Holstein. Las señales inerciales se utilizaron para detectar cambios en el comportamiento físico y así permitir la detección temprana de celo (Anexo [B](#)). Las señales acústicas se utilizaron para el monitoreo del comportamiento alimentario (Anexo [C](#)). La base de datos con señales inerciales tiene una duración más acotada y fue utilizada para experimentos preliminares mientras que la base de datos con señales acústicas contiene registros de varios días de duración y fue utilizada para experimentos exhaustivos.

La detección precisa y confiable de celo en vacas lecheras es esencial para el control de la reproducción y, por lo tanto, para mantener la producción diaria de leche. Los métodos clásicos de detección involucran la identificación visual en busca de ciertos signos que manifiesta el animal. Esta es una tarea prácticamente imposible en grandes rodeos, más aún en el monitoreo simultáneo de muchos animales. Por ello se han propuesto diversas técnicas automáticas para la detección, aunque sólo los métodos más costosos o invasivos han logrado resultados aceptables. En esta tesis se propone un método para la detección de celo por medios no invasivos, basado en el reconocedor de actividades (Anexos [B](#) y [D](#)). Para este problema cada registro se analiza por tramos y la salida del reconocedor es una etiqueta para cada ventana.

Al igual que en el reconocedor de actividades humanas, las señales de entrada fueron las componentes de la aceleración, en conjunto con la magnitud de la misma. Éstas fueron capturadas con un dispositivo ubicado en el collar del animal, posición que pretende capturar los movimientos más importantes del cuerpo. En el preprocesamiento se utilizó un filtro pasa altos para analizar sólo las variaciones de la señal relacionadas a los movimientos más rápidos. El análisis por tramos de la señal se realizó mediante una ventana deslizante, de las cuales se extrajeron las características. La influencia del tamaño de la ventana fue uno de los parámetros analizados (entre 20 y 100 s).

Por otro lado, el monitoreo del comportamiento alimentario en vacas requiere que este se realice a largo de todo el día. Las dos actividades principales, el pastoreo y la rumia, ocupan la mayor parte del día. El pastoreo puede llegar a ocupar entre un 25 % y un 50 % del día mientras que la rumia puede ocupar entre un 15 % y un 40 % del mismo. Tanto la rumia como el pastoreo presentan eventos masticatorios con una marcada regularidad, de aproximadamente 1 evento por segundo. A su vez, la rumia presenta fases periódicas de masticación interrumpidas por breves

pausas, mientras que el pastoreo no posee un patrón predefinido. Utilizando esta información del dominio se ha propuesto y evaluado un reconocedor que puede procesar señales de muy larga duración (días).

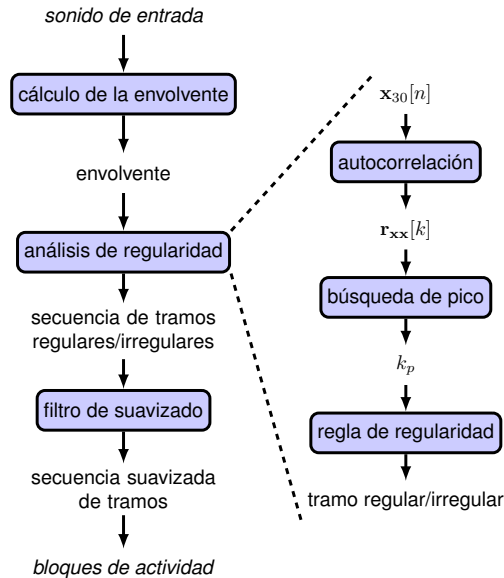


Figura 5: Fases de análisis de la señal de audio para obtener los bloques de actividad.

La entrada al reconocedor propuesto fueron señales acústicas capturadas con un micrófono en la frente del animal. El micrófono apuntó hacia adentro para registrar con mayor intensidad los sonidos masticatorios (que se transmiten a través de los huesos de la cabeza del animal) y al mismo tiempo atenuar los sonidos y ruidos del entorno. Al igual que las señales de aceleración éstas son señales que varían en el tiempo y que poseen información indirecta de la actividad subyacente. La actividad que está realizando el animal debió inferir a partir de los sonidos que registra el dispositivo.

La etapa de preprocesamiento se encarga de obtener una versión muestreada a menor frecuencia de la envolvente de la señal de audio original (Figura 5). Este preprocesamiento disminuye en gran medida la carga computacional de las etapas siguientes que trabajan sobre la envolvente. En segundo lugar, las señales se analizan de forma secuencial mediante tramos de 30 s de la envolvente  $x_{30}[n]$  (Figura 5). Este tamaño brinda robustez al sistema y una granularidad más que suficiente para el análisis de señales de horas de duración. En cada uno de estos tramos se busca evidencia de que existan eventos con una regularidad cercana a 1

s, utilizando la autocorrelación de la envolvente,

$$r_{xx}[k] = \sum_{n=k}^{N_{30}-1} x_{30}[n]x_{30}[n-k],$$

donde  $k \geq 0$  es el desplazamiento y  $N_{30}$  es el número de muestras en el tramo. Un tramo será marcado como regular si el pico  $k_p$  está en el intervalo de regularidad  $L_{\text{reg}}$ , cuando es buscado en un intervalo más amplio  $L_{\text{peak}}$ ,

$$k_p = \arg \max_{k \in L_{\text{peak}}} \{r_{xx}[k]\}.$$

A partir de los tramos identificados como regulares se construyen bloques continuos de actividad, los cuales son el resultado de la segmentación de cada registro analizado y serán clasificados en una etapa posterior.

## 4.2. Extracción de características

El objetivo de la extracción de características es describir cada tramo o bloque de la señal de forma que pueda ser clasificada. Idealmente, la extracción daría origen a una representación que haga trivial el trabajo del clasificador, donde los vectores sean similares para objetos en la misma categoría y diferentes para objetos en diferentes categorías. Esto lleva a la idea de características discriminativas, que a la vez sean invariantes a transformaciones irrelevantes de la señal de entrada. Por ejemplo, que la frecuencia de los pasos al caminar varíe entre una persona y otra no debería provocar grandes cambios en las características para reconocer la actividad. Incluso, una misma persona puede cambiar la frecuencia de su caminar en distintos momentos. En cambio la frecuencia de los pasos puede ayudar a distinguir entre dos actividades de ritmos claramente distintos, como caminar y correr. También, se deberían buscar características que sean independientes a traslaciones temporales o modificaciones de la duración de la actividad. Caminar por 5 o 10 minutos sigue siendo la misma actividad, y las características no deberían verse modificadas por tal variación de la duración. Dado que la extracción de características es muy dependiente del problema, se propone en cada aplicación la elección de las características, en base a los antecedentes en el área y el conocimiento del dominio.

### Reconocimiento de actividades humanas

Para el reconocimiento de actividades humanas se propone una representación que brinde información de la dinámica del cuerpo del individuo. Dicha información está relacionada a la actividad que esté desarrollando el individuo. Consideremos la hipótesis de que la señal de aceleración  $a(t)$  es la salida de un sistema convolutivo.

La excitación del sistema,  $m(t)$ , puede asociarse al ritmo del movimiento, originado por las fuerzas de la actividad muscular y las interacciones del cuerpo con el exterior. Entonces, la dinámica del cuerpo se puede modelar en la respuesta al impulso  $h(t)$ , y dependerá de la actividad que esté desarrollando el individuo. Como el interés es recuperar información acerca de la actividad desarrollada, el análisis a utilizar debe ser capaz de aislar  $h(t)$ . Sin embargo, tanto la excitación como la respuesta al impulso son desconocidas, entonces para recuperar  $h(t)$  sería necesario realizar una deconvolución ciega.

El análisis homomórfico fue desarrollado como un método general para separar señales que estén combinadas de forma no aditiva [73]. El objetivo es convertir esa combinación en una combinación lineal, donde las técnicas de análisis clásicas se puedan aplicar más directamente. Por ejemplo, cuando dos señales están convolucionadas en el dominio temporal

$$a(t) = m(t) * h(t),$$

sus transformadas de Fourier se multiplican en el dominio frecuencial,

$$A(f) = M(f) \cdot H(f),$$

lo cual puede transformarse en una suma aplicando un logaritmo apropiadamente definido,

$$\log |A(f)| = \log |M(f)| + \log |H(f)|.$$

En este punto  $\log |M(f)|$  y  $\log |H(f)|$  se encuentran combinados aditivamente. Aplicando la transformada inversa de Fourier al  $\log |A(f)|$  se puede obtener una nueva representación temporal

$$c(q) = \mathcal{F}^{-1} \{ \log |A(f)| \}.$$

Esta representación es la que se conoce como cepstrum y su dominio son las frecuencias. En las primeras componentes de  $c(q)$  se codifican las variaciones más lentas de  $\log |A(f)|$ . Las demás componentes de  $c(q)$  codifican las variaciones más rápidas de  $\log |A(f)|$ . Generalmente, éstas se encuentran separadas en el cepstrum, y pueden relacionarse con la respuesta al impulso  $h(t)$  y la excitación del sistema  $m(t)$ , respectivamente. Entonces, tomando los primeros coeficientes de  $c(q)$  se puede recuperar la información deseada sobre la dinámica del cuerpo o movimiento.

En el dominio del tiempo las señales de aceleración muestran un comportamiento cuasi-periódico, como se observa en la Figura 6a. Para este ejemplo, el periodo es de  $\sim 0.5$  s, lo cual es un ritmo esperado para una actividad humana cotidiana. Esta señal brinda parte de la información de la señal de excitación  $m(t)$ , pero sin embargo permanece oculta la respuesta al impulso del sistema  $h(t)$ . En el dominio de las frecuencias, este periodo se puede ver como un pico en la frecuencia

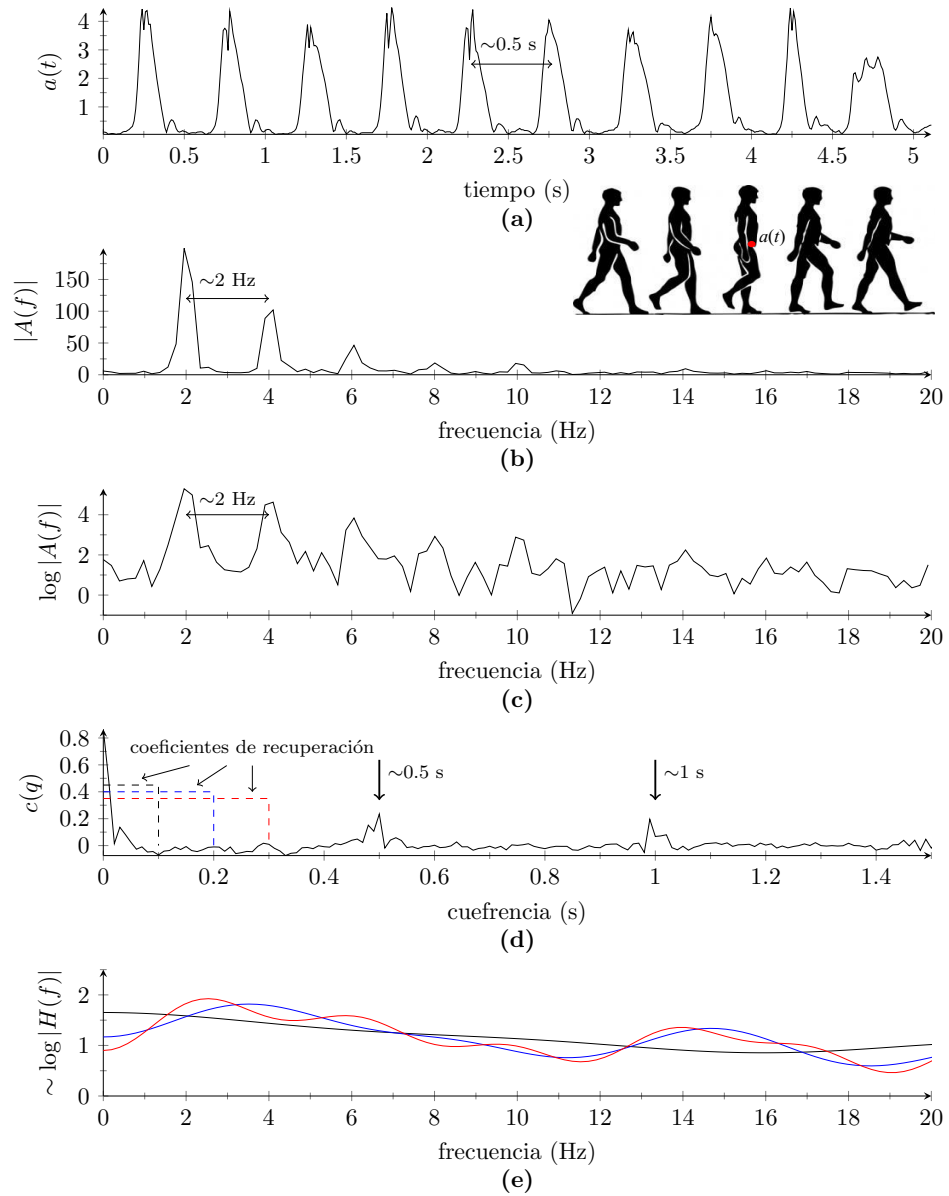


Figura 6: Obtención de los coeficientes cepstrales. (a) Señal de aceleración cuasi-periódica  $a(t)$ ; (b) espectro de  $a(t)$ ; (c) logaritmo del espectro; (d) coeficientes cepstrales; (e) aproximaciones de  $\log |H(f)|$  recuperadas.

fundamental  $\sim 2$  Hz y en sus correspondientes armónicos (Figura 6b). Estos son los mayores aportes, quedando enmascarada la información restante de la respuesta al impulso del sistema. La Figura 6c muestra la señal que se obtiene luego de aplicar el logaritmo a  $|A(f)|$ . A continuación, al aplicar la transformada inversa de Fourier se obtiene el cepstrum  $c(q)$  (Figura 2d). Los picos relacionados a la señal de excitación todavía están presentes en  $\sim 0.5$  s y  $\sim 1.0$  s pero ahora ya no se solapan con la representación cepstral de la respuesta al impulso.

En la etapa de extracción de características se consideraron tres tipos de características: los coeficientes cepstrales (CC), los atributos temporales (AT) y el período fundamental (PF). Los CCs refieren a los primeros coeficientes de la representación cepstral. Los AT fueron 5 valores estadísticos calculados a partir de  $a(t)$ : desvío estándar, energía, máximo, mínimo, y amplitud. El PF consiste en un único valor que representa el ritmo del movimiento, extraído de  $a(t)$  utilizando la autocorrelación sin sesgo [74].

Dado que  $a(t)$  puede representar a  $a_x(t)$ ,  $a_y(t)$ ,  $a_z(t)$  o  $|\mathbf{a}(t)|$ , cada una de estas señales puede dar origen a su propio conjunto de características. Entonces, el vector de características final podrá estar integrado por diferentes combinaciones. Se analizaron un total de 32 combinaciones de características obtenidas de distintas señales de entrada. En el caso de los coeficientes cepstrales también se evaluó la cantidad de coeficientes a tomar en la representación.

## Reconocimiento de actividades en vacas lecheras

Para la detección de celo, los vectores de características fueron conformados con los cinco atributos temporales (AT) utilizados para el reconocimiento de actividades humanas, obtenidos de las componentes y la magnitud de la señal de aceleración. Las características fueron extraídas luego de aplicar el filtro pasa altos en la etapa de preprocesamiento, y sin aplicarlo. Además, se propuso realizar una selección de características para eliminar la redundancia entre ellas.

Para la clasificación de las actividades alimentarias, el primer paso es caracterizar los bloques de actividad identificados durante la segmentación. Dentro de los bloques de actividad, se analizan las señales de sonido mediante tramos de 1 s,  $x_1[n]$ . Estos permiten detectar las interrupciones relacionadas a la regurgitación del bolo alimenticio durante la rumia (indicados con flechas rojas en la Figura 7b). Las interrupciones podrían inferirse a partir de cambios en la amplitud de la señal sonora. Sin embargo, no es confiable la detección debido a las variaciones en la amplitud a lo largo de las sesiones, los distintos micrófonos, dispositivos de grabación y animales. En experimentos preliminares se encontró que una medida relacionada a la energía da mejores resultados para la detección. Por ello, para el reconocimiento de actividades alimentarias, se caracteriza cada tramo mediante  $e[j] = \log(\|\mathbf{x}_1\|_2)$ . El centro de cada tramo analizado se indica con  $j$ . Por ejemplo,

la línea sólida en la Figura 7c muestra la energía  $e[j]$  correspondiente al sonido registrado durante la rumia (Figura 7b). El algoritmo propuesto consiste en detectar las caídas repentinas en  $e[j]$  y caracterizar su regularidad. De esta manera, un bloque es caracterizado por dos atributos: el intervalo promedio entre interrupciones y la tasa de interrupciones por minuto.

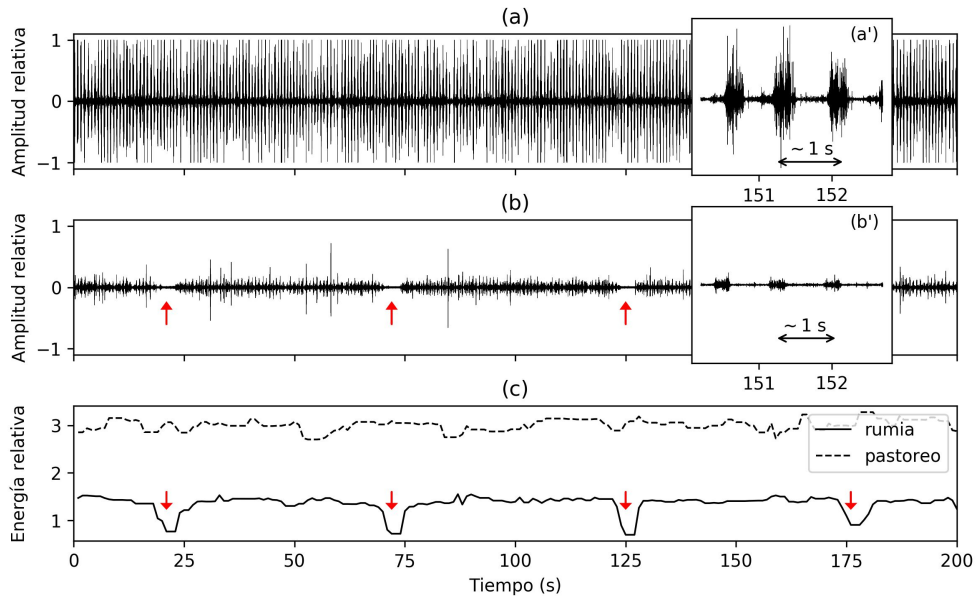


Figura 7: Sonidos registrados durante a) el pastoreo y b) la rumia. c) Energía de la señal sonora durante el pastoreo (línea de puntos) y la rumia (línea continua). Se indican con flechas rojas las interrupciones relacionadas a la regurgitaciones del bolo alimenticio.

### 4.3. Clasificación y posprocesamiento

La misión del clasificador es tomar el vector de características y asignarlo a la categoría o la clase que le corresponde. En ocasiones, en vez de asignar directamente una de las clases se puede optar por modelos que asignen una probabilidad para cada clase. El nivel de dificultad de la clasificación depende en gran medida de la variabilidad de las características entre objetos de la misma clase y la diferencia relativa entre objetos de diferentes clases. En esta tesis utilizamos aquellos métodos que mejor se han desempeñado de acuerdo a los antecedentes. La clasificación se realizó considerando cada ventana o bloque por separado.

Completada la etapa de clasificación, la salida del reconocedor puede ser modificada al considerar información del dominio que no está embebida en los vectores

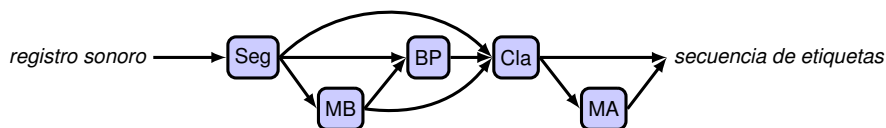


de características o las señales originales. Por ejemplo, como resultado de la clasificación podrían surgir actividades con una duración más breve de lo esperado, o una secuencia de actividades que no es admisible por el ámbito donde se realiza el reconocimiento. En esta etapa de pos procesamiento se incorpora conocimiento del dominio para imponer este tipo de restricciones y que la salida final del reconocedor se acerque lo máximo posible a la realidad. De ser posible, esta información del dominio puede ser suministrada al clasificador como una característica más para realizar la clasificación.

En el caso de los registros de actividades humanas previamente segmentadas (SCUT-NAA), para la clasificación se utilizaron máquinas de soporte vectorial, perceptrones multicapa, árboles aleatorios y clasificadores Bayesianos ingenuos. En la etapa de pos procesamiento se consideran las etiquetas asignadas a cada ventana para definir la etiqueta final de cada registro. En el caso de los registros con secuencias de actividades humanas (HASC), el clasificador elegido fue el perceptrón multicapa. El perceptrón multicapa es versátil y posee la particularidad de brindar a su salida la probabilidad por clase. En las secuencias de etiquetas esto permitió definir un umbral mínimo de confianza para asignar la etiqueta final. Al igual que en los resultados de referencia [24], se utilizaron los registros segmentados por actividad para el entrenamiento de los modelos del clasificador. La evaluación se realizó sobre señales que se corresponden a secuencias continuas de actividades, diferentes a los registros del entrenamiento. En ambos problemas, en el entrenamiento se consideró una búsqueda de grilla para definir los parámetros óptimos de los clasificadores (por ejemplo, la cantidad de nodos en la capa oculta).

Para la etapa de clasificación en la detección de celo se consideraron tres algoritmos clásicos: el perceptrón multicapa, la máquina de soporte vectorial y el árbol de decisión. Al igual que en el reconocimiento de actividades humanas, la clasificación se realizó por ventana. En cuanto a la clasificación de actividades alimentarias en vacas lecheras, ésta se realizó a partir de la caracterización de las interrupciones detectadas. En este caso se utilizaron reglas heurísticas para la clasificación, las cuales fueron ajustadas con un subconjunto de registros de la base de datos. Este subconjunto de registros no fue utilizado para evaluar el desempeño del algoritmo propuesto. Además de la estrategia principal de segmentación (Seg) y clasificación (Cla) antes mencionada, también se analizaron otras variantes del algoritmo, incluyendo etapas de procesamiento intermedias o posteriores (Tabla 5). Dos de estas etapas son la fusión de los pequeños bloques de inactividad antes (MB) o después (MA) de la clasificación. Estas etapas están basadas en el conocimiento del dominio y principalmente buscan reducir la cantidad de bloques cortos de actividad o inactividad, que no coincide con el comportamiento natural de los animales. Por último, se propuso una etapa intermedia para la partición de bloques largos. Para ello, la partición se realizó mediante la detección de cambios en la energía de la

Tabla 5: Variantes del algoritmo para el reconocimiento de actividades alimentarias.




---

RAFAR:	$\text{registro sonoro} \rightarrow \text{Seg} \rightarrow \text{Cla} \rightarrow \text{secuencia de etiquetas}$
RAFAR-MB:	$\text{registro sonoro} \rightarrow \text{Seg} \rightarrow \text{MB} \rightarrow \text{Cla} \rightarrow \text{secuencia de etiquetas}$
RAFAR-MA:	$\text{registro sonoro} \rightarrow \text{Seg} \rightarrow \text{Cla} \rightarrow \text{MA} \rightarrow \text{secuencia de etiquetas}$
RAFAR-BP:	$\text{registro sonoro} \rightarrow \text{Seg} \rightarrow \text{BP} \rightarrow \text{Cla} \rightarrow \text{secuencia de etiquetas}$
RAFAR-BPMA:	$\text{registro sonoro} \rightarrow \text{Seg} \rightarrow \text{BP} \rightarrow \text{Cla} \rightarrow \text{MA} \rightarrow \text{secuencia de etiquetas}$
RAFAR-MBBP:	$\text{registro sonoro} \rightarrow \text{Seg} \rightarrow \text{MB} \rightarrow \text{BP} \rightarrow \text{Cla} \rightarrow \text{secuencia de etiquetas}$

---

señal sonora (BP).

#### 4.4. Validación y medidas de desempeño

Evaluar el desempeño de un reconocedor requiere de la utilización de diversos análisis estadísticos y métodos de validación. En términos generales, en esta tesis se propone utilizar validación cruzada, cuidando que cada registro sea asignado completamente al conjunto de entrenamiento o al de validación. Por ejemplo, cuando se utiliza validación cruzada dejando un sujeto afuera, las señales capturadas sobre dicho sujeto sólo son consideradas para la validación y no son parte del conjunto utilizado para el entrenamiento del reconocedor. En la Figura 8 se brinda un esquema simplificado de validación cruzada con 5 particiones. En la primera prueba de este esquema, la primera parte de los datos se reserva para la validación, mientras que los datos restantes se utilizan para seleccionar los mejores parámetros del reconocedor. En las pruebas siguientes se reservan otras porciones de los datos para la validación, continuando hasta que todos los datos hayan sido considerados. Salvo que se indique lo contrario, en los resultados se reportará el promedio de las medidas de desempeño obtenidas a lo largo de todas las pruebas de la validación.

Cuando el problema de clasificación es estándar, a cada vector de entrada se le asigna una de las clases posibles. En este caso, las definiciones de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN) están claramente establecidas. Entonces, el desempeño de un reconocedor puede evaluarse utilizando medidas estándares tales como la tasa de acierto, la precisión y la sensibilidad [67]. El reconocimiento de actividades en registros segmentados por actividad se corresponde con un problema de clasificación estándar. Por lo que

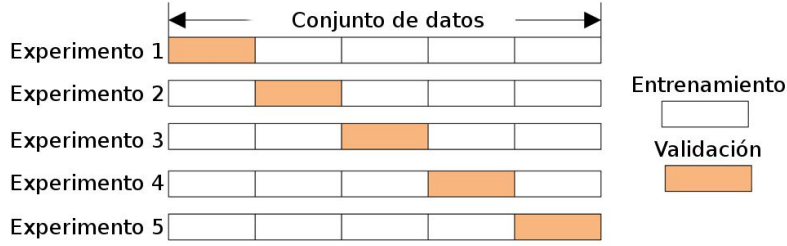


Figura 8: Esquema simplificado de la validación cruzada.

es posible utilizar la tasa de acierto

$$a = \frac{\sum_i VP_i + VN_i}{\sum_i VP_i + VN_i + FP_i + FN_i},$$

la precisión

$$p = \frac{\sum_i VP_i}{\sum_i VP_i + FN_i}$$

y la sensibilidad

$$s = \frac{\sum_i VP_i}{\sum_i VP_i + FP_i},$$

para cada clase  $i$ . Cuando las tasas de acierto son altas es más útil medir la reducción relativa del error:  $\delta = (\epsilon_B - \epsilon_A)/\epsilon_A$ , donde  $A$  es el sistema de referencia para dos reconocedores con errores absolutos  $\epsilon_A$  y  $\epsilon_B$ .

En el reconocimiento continuo de actividades se debe comparar una secuencia de referencia con una secuencia reconocida. Los bloques de actividad de una secuencia y otra es muy posible que ya no tengan una relación uno a uno, como sí la hay en un problema de clasificación estándar. Identificar los tipos de errores que comete el reconocedor permiten comprender cómo está funcionando y permite orientar los esfuerzos para superarlos. En la Figura 9 se dan las posibles situaciones en la comparación de dos secuencias de actividad, cuando hay una relación uno a uno Figura 9(a) y cuando no la hay Figura 9(b). Las dos primeras situaciones en la parte (a) se corresponden a un bloque de actividad perfectamente detectado y un período de inactividad correctamente reconocido. Los dos siguientes se corresponden a bloques correctamente identificados, aunque sobreestimando su longitud en el primero y subestimando la misma en el segundo. Estos tipos de errores pueden darse al inicio de un bloque, al final, o a ambos lados (como en los ejemplos). Por último, las dos situaciones finales se corresponden a la completa eliminación de un bloque que estaba presente en la secuencia de referencia y, al agregado de un bloque en la secuencia reconocida, respectivamente.

Las situaciones en la Figura 9(b) se corresponden con relaciones más complejas entre los bloques de la secuencia de referencia y la reconocida. Estas son las si-

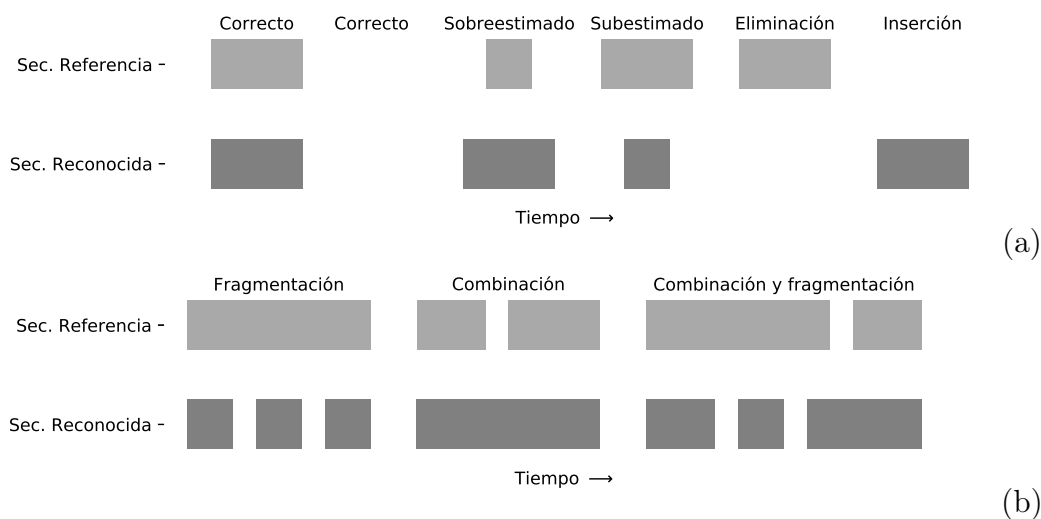


Figura 9: Posibles situaciones de la comparación entre la secuencia de referencia y reconocida, (a) conservando la relación uno a uno entre bloques o (b) sin relación uno a uno entre bloques.

tuaciones que no contemplan las medidas de clasificación tradicionales. El primer bloque de la referencia es reconocido de forma fragmentada con tres bloques en la secuencia reconocida. El segundo y tercer bloque de la referencia son reconocidos por un solo bloque que los combina a ambos. Finalmente, los dos últimos bloques de la referencia son reconocidos cometiendo fragmentación y combinación de forma simultánea. Otras situaciones posibles que no se muestran en la figura involucran la sobreestimación y/o subestimación simultánea junto con otros errores (bloques fragmentados, combinados, combinados y fragmentados).

Esta complejidad en la comparación provoca que las definiciones de VP, VN, FP y FN ya no sean útiles. Redefinir el problema para que sí puedan aplicarse las medidas estándares puede llevar a resultados confusos o incompletos. Por ejemplo, para recuperar la relación uno a uno y las medidas tradicionales, se podría analizar el reconocimiento por ventanas, el cual brinda detalle temporal. Sin embargo, de esta forma los errores que podrían considerarse graves (como las inserciones o eliminaciones de bloques) no se distinguen de otros menores, como los bloques parcialmente detectados. Dependiendo del tamaño de las ventanas se tendrá mayor o menor detalle en temporal, aunque siempre en el corto plazo. Por ello es de interés considerar medidas multidimensionales del reconocimiento, en escalas de corto y largo plazo.

En este sentido, para el análisis del reconocimiento continuo, se trabajó a partir de las medidas propuestas en [69], utilizando una nueva representación gráfica multidimensional para el análisis. Éstas han sido poco usadas en el reconocimiento

Tabla 6: Subcategorías de los segmentos identificados inicialmente como VP, FP, VN o FN.

Subcategoría	Detalle
C (correcto)	un segmento identificado como VP
I (inserción)	un segmento FP que se corresponde con una inserción aislada en la secuencia reconocida
M (combinación)	un segmento FP rodeado por segmentos VP de un bloque combinado
Ua (subcompleto)	un segmento FP que ocurre al principio de un bloque parcialmente identificado
Uz (subcompleto)	un segmento FP que ocurre al final de un bloque parcialmente identificado
D (eliminación)	un segmento FN que se corresponde con una eliminación aislada en la secuencia reconocida
F (fragmentación)	un segmento FN rodeado por segmentos VP de un bloque fragmentado
Oa (sobrecompleto)	un segmento FN que ocurre al principio de un bloque parcialmente detectado
Oz (sobrecompleto)	un segmento FN que ocurre al final de un bloque parcialmente detectado

de actividades humanas, y nunca utilizadas en el reconocimiento de actividades animales. La comparación de dos secuencias binarias se basa en la noción de segmento. En este contexto, un segmento está definido como la porción más larga de un bloque en la cual la comparación entre la secuencia de referencia y la reconocida puede realizarse sin ambigüedad (segmentos claramente definidos como VP, FP, VN y FN). Los segmentos no tienen una longitud fija y son derivados de la comparación entre la secuencia de referencia y la reconocida, un cambio en cualquiera de las dos secuencias marca el límite de un segmento. Este aspecto marca una clara diferencia de los bloques y las ventanas, que pueden definirse a partir de una sola secuencia. Los segmentos que resultan ser VN no ven modificadas sus etiquetas. En cambio, los segmentos considerados como VP, FP o FN son reasignados a nuevas subcategorías (Tabla 6). La categorización final de las ventanas se obtiene a partir de los segmentos y un bloque se etiqueta a partir de todos los segmentos que se solapan con él.

## 5. Resultados

En las secciones siguientes se describen los resultados más importantes al aplicar los reconocedores propuestos para cada problema. Por un lado, se presentan los resultados en el reconocimiento de actividades humanas a partir de señales de aceleración. Por otro lado, se presentan los resultados en el reconocimiento de actividades en vacas lecheras, a partir de señales de aceleración y señales acústicas.

Tabla 7: Desempeño global comparado mediante el error de reconocimiento.

Actividad	Mejor clasificador conocido [26]	Sistema propuesto	Reducción relativa del error (%)
Caminar de a pasos (s)	13.64	<b>2.27</b>	83.36
Saltar (j)	2.33	<b>0.00</b>	100.00
Bicicleta (b)	16.67	<b>0.00</b>	100.00
Caminar (w)	<b>22.73</b>	25.00	–
Caminar hacia atrás (wb)	18.18	<b>9.30</b>	48.84
Caminar rápido (wq)	<b>18.18</b>	23.81	–
Correr (r)	2.27	<b>0.00</b>	100.00
Relajarse (re)	4.55	<b>2.27</b>	50.1
Bajar las escaleras (d)	18.18	<b>13.64</b>	24.97
Subir las escaleras (u)	15.91	<b>9.30</b>	41.55
Global	13.18	<b>8.79</b>	33.31

## 5.1. Reconocimiento de actividades humanas

Para el reconocimiento de las actividades humanas, primero se trabajó con señales de aceleración donde cada registro estuviera segmentado por actividad. En estos experimentos, principalmente se trabajó en la etapa de extracción de características del reconocedor. Las actividades humanas reconocidas fueron aquellas que desarrolla una persona en su vida diaria, tales como caminar, permanecer parado, correr, subir y bajar escaleras. Los resultados completos de estos experimentos fueron publicados en el artículo del Anexo [A](#) (aquí destacamos los más importantes). Más recientemente se realizaron experimentos extendiendo el reconocimiento a registros que consisten en una secuencia de actividades. En este caso el desafío adicional es identificar los momentos que se corresponden con cada actividad, además de distinguir cual es la actividad.

### Registros segmentados por actividad

Para la evaluación se realizó validación cruzada dejando un sujeto afuera, donde cada reconocedor fue construido de forma independiente al sujeto de evaluación. Esta construcción permitió evaluar al reconocedor en el escenario de no estar ajustado al sujeto sobre el cual se realiza el reconocimiento. Como medida de desempeño se utilizó la tasa de acierto de actividades. También, se consideraron la precisión, la sensibilidad y la matriz de confusión para analizar los resultados y realizar comparaciones entre las distintas actividades a reconocer. Dado que las tasas de acierto se encontraban cercanas al 100 % también se utilizó la reducción relativa del error para comparar el desempeño entre los reconocedores.

Los primeros experimentos evaluaron distintos tamaños de la representación

Tabla 8: Matriz de confusión para el vector de características  $F_{24}$  con 70 CCs.

Actividad	s	j	b	w+wq	wb	r	re	d	u
Caminar de a pasos (s)	43		1						
Saltar (j)		43							
Bicicleta (b)			30						
Caminar hacia adelante (w+wq)				80	1			2	3
Caminar hacia atrás (wb)				3	39				1
Correr (r)						44			
Relajado (re)			1				43		
Bajar las escaleras (d)					1	3		38	2
Subir las escaleras (u)				1	1			3	38

cepstral, a partir de las componentes y la magnitud de la señal. Luego se analizó la combinación de la nueva representación junto a las otras características. Los resultados indicaron que la composición más adecuada para el vector de características es aquella que combina lo mejor de cada característica propuesta. Posee una representación cepstral amplia (70 CCs de  $|\mathbf{a}(t)|$ ), un compendio completo de los atributos temporales (AT extraídos de la magnitud y las componentes), y el ritmo al que se ejecutó la actividad (PF de  $|\mathbf{a}(t)|$ ). Las comparaciones del reconocedor propuesto con el estado del arte revelaron que la combinación de características de distinto tipo (cepstrales, temporales y ritmo) resultó ser más robusta que sólo considerar las características espectrales (Tabla 7). En promedio, el sistema propuesto redujo el error de reconocimiento en un 33% comparado con el estado del arte. La ventaja de las características cepstrales es que comprimen la información espectral en unos pocos coeficientes, a la vez que retienen el poder discriminativo. Por otro lado, los resultados mostraron que al modelar conjuntamente las actividades de caminar (w) y caminar rápido (wq) se puede reducir en gran medida las posibles confusiones (Tabla 8). Además, se encontró que puede alcanzarse un alto desempeño aún si se considera sólo a la señal de la magnitud de la aceleración como entrada. En este caso se obtuvo una tasa de reconocimiento promedio del 90.97%, cercano al 91.21% de la mejor combinación de características, y ambos por encima del 86.82% del estado del arte.

### Secuencias de actividades

El desempeño obtenido para el reconocimiento de actividades continuas se resume en la Tabla 9. En este caso el desempeño se presenta según la medida-F calculada por ventanas. Los resultados muestran que las dos actividades peor reconocidas son bajar y subir escaleras. Aún así, los reconocedores propuestos superan a los resultados de referencia para estas actividades. La actividad caminar fue aceptablemente reconocida, aunque se observó cierta confusión con las actividades

Tabla 9: Comparación de las distintas variantes de la etapa de clasificación. Los valores de las celdas corresponden a la medida-F1, calculada por ventana.

Actividad	Base	Suavizado	Anulación	Anulación y suavizado	Referencia <span style="border: 1px solid green; padding: 0 2px;">24</span>
Trotar	0.85	<b>0.88</b>	0.87	0.85	<b>0.88</b>
Trotar dando saltos	0.84	0.86	<b>0.87</b>	0.82	0.83
Permanecer de pie	0.90	0.87	0.90	0.81	<b>0.95</b>
Bajar escaleras	0.56	0.64	0.59	<b>0.65</b>	0.48
Subir escaleras	0.62	<b>0.67</b>	0.61	0.66	0.58
Caminar	0.74	<b>0.77</b>	0.72	0.72	0.64
Total	0.76	<b>0.79</b>	0.77	0.75	0.73

de subir y bajar escaleras. Estas confusiones coinciden con los resultados obtenidos en la base de datos SCUT-NAA. La similitud tanto de los movimientos como de sus cadencias parecieran trasladarse a las características utilizadas. Por último, trotar y trotar dando saltos, fueron las actividades mejor reconocidas.

Al aplicar la etapa de suavizado a las secuencias de la clasificación base, se mejoró el reconocimiento en todas las actividades excepto para permanecer de pie. Este suavizado redujo la cantidad de ventanas que fueron aisladamente clasificadas como de una actividad. En forma global el reconocimiento supera a la clasificación base. La aplicación de la etapa de anulación por umbral también mejora la clasificación base. Esto se debe a la reducción de los falsos positivos, dado que las ventanas en las que el clasificador no tiene alta confianza se descarta la etiqueta. La anulación y el suavizado provee los mejores resultados para bajar escaleras pero en las restantes actividades fue superado por las otras alternativas de pos proceso.

## 5.2. Reconocimiento de actividades en vacas lecheras

En esta sección se resumen los resultados en vacas lecheras para la detección de celo y el reconocimiento de actividades alimentarias. La descripción detallada de los experimentos llevados a cabo y sus resultados se pueden encontrar en los Anexos B y C, respectivamente.

En cuanto a los experimentos para la detección de celo, se utilizaron medidas de desempeño basadas en ventanas y se eligió la opción de dejar un registro afuera para la validación cruzada, dado que la base de datos no era muy extensa. Además de evaluar el tamaño de la ventana de análisis, se consideró realizar una selección de características. Los mejores resultados se obtuvieron con una ventana de análisis de 40 s, considerando todas las características y sin aplicar el filtrado pasa altos en la etapa de preprocesamiento. Los resultados preliminares fueron muy buenos, con tasas de detección superiores a 90 %, siendo el perceptrón multicapa el clasificador



Tabla 10: Resumen de la segmentación y la clasificación de actividades alimentarias. La medida-F basada en ventanas y en bloques se promedió a lo largo de todos los registros analizados (desvío estándar).

	Segmentación de actividades		Clasificación de rumia		Clasificación de pastoreo	
	Por ventanas	Por bloques	Por ventanas	Por bloques	Por ventanas	Por bloques
RAFAR	0.943 ( $\pm 0.097$ )	0.612 ( $\pm 0.347$ )	0.780 ( $\pm 0.208$ )	0.703 ( $\pm 0.288$ )	0.849 ( $\pm 0.191$ )	0.770 ( $\pm 0.291$ )
RAFAR-MB	<b>0.962</b> ( $\pm 0.057$ )	<b>0.715</b> ( $\pm 0.303$ )	0.778 ( $\pm 0.176$ )	0.818 ( $\pm 0.231$ )	0.878 ( $\pm 0.141$ )	0.829 ( $\pm 0.233$ )
RAFAR-MA	0.956 ( $\pm 0.068$ )	0.688 ( $\pm 0.322$ )	0.789 ( $\pm 0.206$ )	0.791 ( $\pm 0.221$ )	0.854 ( $\pm 0.180$ )	0.787 ( $\pm 0.269$ )
RAFAR-BP	0.943 ( $\pm 0.097$ )	0.612 ( $\pm 0.347$ )	0.836 ( $\pm 0.178$ )	0.719 ( $\pm 0.274$ )	0.882 ( $\pm 0.157$ )	0.770 ( $\pm 0.288$ )
RAFAR-BPMA	0.956 ( $\pm 0.068$ )	0.688 ( $\pm 0.322$ )	0.844 ( $\pm 0.172$ )	0.813 ( $\pm 0.227$ )	0.885 ( $\pm 0.151$ )	0.796 ( $\pm 0.274$ )
RAFAR-MBBP	<b>0.962</b> ( $\pm 0.057$ )	<b>0.715</b> ( $\pm 0.303$ )	<b>0.891</b> ( $\pm 0.125$ )	<b>0.873</b> ( $\pm 0.191$ )	<b>0.935</b> ( $\pm 0.114$ )	<b>0.852</b> ( $\pm 0.225$ )

de mejor desempeño para el reconocedor propuesto.

Respecto al reconocimiento de actividades alimentarias, los resultados fueron altamente satisfactorios. Para analizar el desempeño de las variantes resultó de gran utilidad las métricas por ventanas y por bloques. Esta visión dual, de corto y largo plazo, resultó en varios beneficios, tanto para el desarrollo del reconocedor como para el análisis comparativo. Por un lado, permitió identificar los tipos de errores cometidos por cada variante y así facilitar la toma de decisiones para proponer mejoras. Por otro lado, facilitó la comparación multidimensional entre las variantes, análisis que no ha sido aplicado en el reconocimiento de actividades animales. También se tuvo en cuenta una medida práctica: la estimación de tiempos de pastoreo y rumia. En particular, la estimación de tiempos de rumia permitió comparar el sistema propuesto con un sistema comercial que actualmente se utiliza en diversos establecimientos a lo largo del mundo.

En la Tabla [10](#) se resumen los resultados de la segmentación y la clasificación de actividades alimentarias. Los resultados corresponden a la medida-F basada en ventanas y en bloques, al realizar el promedio a lo largo de todos los registros analizados. Los resultados calculados por ventanas fueron excelentes para la segmentación (arriba de 0.94). Esto quiere decir que el algoritmo propuesto fue muy preciso para discriminar entre actividades alimentarias y otras actividades. Las medidas calculadas por bloques alcanzaron valores de hasta 0.715 para la segmentación. Esto indica que la mayoría de los bloques de actividad fueron correctamente identificados. Entre los bloques que no fueron detectados correctamente,

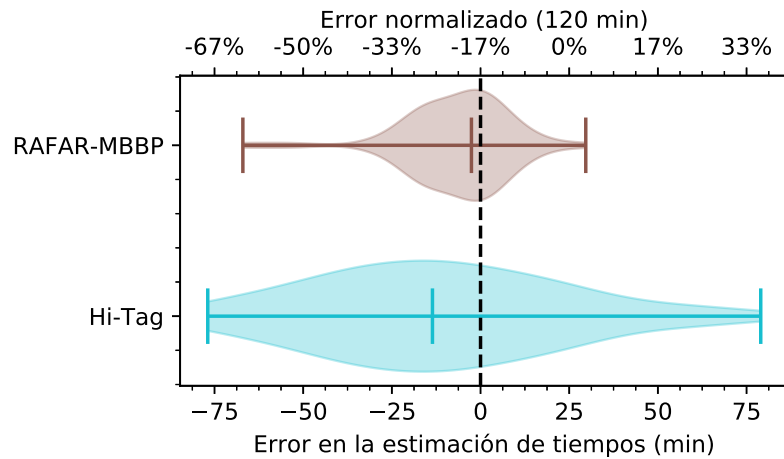


Figura 10: Error en la estimación de los tiempos de rumia para RAFAR-MBBP y el sistema Hi-Tag. El eje superior está normalizado por la longitud de los intervalos analizados (2 horas).

el análisis detallado de las secuencias reconocidas reveló que gran parte de ellos fueron parcialmente detectados. La mejor segmentación se obtuvo con las variantes RAFAR-MB y RAFAR-MBBP. En cuanto a la clasificación, las medidas obtenidas por ventanas fueron algo menores si son comparadas con los resultados de la segmentación. Esto indica que algunos de los bloques correctamente segmentados son incorrectamente clasificados. Por el contrario, los resultados por bloques en cuanto a la clasificación son superiores a los que se obtuvieron para la segmentación. Esto es particularmente evidente para RAFAR-MBBP, variante que combina una etapa para eliminar la fragmentación (MB) y una etapa para partir los bloques largos antes de clasificarlos (BPS). El exhaustivo análisis de los resultados demostró que la mejor variante es RAFAR-MBBP y fue la utilizada para la comparación con el sistema comercial Hi-Tag.

Los errores de estimación de los tiempos de rumia se muestran en la Figura 10. El sistema comercial cometió grandes errores de subestimación (hasta -75 min) y sobreestimación (hasta 60 min). Estos valores corresponden a más del 50% de los intervalos considerados. La distribución de los errores arrojó una mediana de -13,55 min. Por el contrario, el sistema propuesto cometió errores mucho menores. Éstos se encontraron acotados entre -25 min de subestimación y 15 min de sobreestimación. En este caso, la distribución de los errores también mostró una tendencia a la subestimación, pero con una mediana mucho menor, de -2,56 min. La explicación del éxito del algoritmo propuesto para estimar la rumia se basa en dos aspectos típicos de la actividad, los cuales son utilizados para el reconocimiento: (i) la

marcada regularidad de las masticaciones y (ii) las interrupciones cuasi-periódicas relacionadas a las degluciones del bolo alimenticio.

## 6. Conclusiones y trabajos futuros

En esta tesis se trabajó en el reconocimiento de actividades humanas y animales y las distintas etapas necesarias para realizar estas tareas. Partiendo desde el esquema general del reconocedor, las distintas etapas fueron adaptadas y mejoradas para cada problema específico. En todas las aplicaciones las señales de entrada, inerciales o acústicas, fueron capturadas con sensores dispuestos sobre el cuerpo. Esta estrategia permitió realizar el monitoreo de las actividades sin restringir el ambiente donde la persona o el animal realicen sus actividades.

El recorrido por las diferentes aplicaciones permitió ahondar en los diferentes desafíos del reconocimiento de actividades y de cada uno de las etapas involucradas. En cuanto a las señales, se encontró que a partir de un adecuado análisis unidimensional se pueden obtener altas tasas de reconocimiento. Esto permite pensar en que las etapas iniciales del reconocedor pueden ser simplificadas y el preprocesamiento de las señales embebido en los mismos dispositivos de captura. Los sistemas basados en señales acústicas se verían muy beneficiados por esta simplificación.

Respecto al análisis por tramos, la utilización de ventanas de tamaño fijo demostró ser versátil para los diferentes problema con las señales inerciales. En el caso de las señales acústicas, además se trabajó en el desarrollo de una etapa de segmentación, que separe los bloques de actividad de otros momentos irrelevantes. Mediante la autocorrelación de la envolvente de la señal de audio se logró detectar eventos masticatorios periódicos. Esta detección permitió delimitar los bloques de actividad para su posterior análisis y clasificación. La cantidad y longitud de las señales analizadas (varias horas) han permitido evaluar exhaustivamente las técnicas desarrolladas. Además de las actividades alimentarias, en estos registros hay largos bloques irrelevantes que el algoritmo ha descartado adecuadamente.

En cuanto a la extracción de características en el reconocimiento de actividades humanas, se propuso una nueva representación para las señales de aceleración: la representación cepstral. Esta brinda una representación más compacta que supera en desempeño a las características típicamente utilizadas. Además se analizaron las mejoras al combinar esta nueva representación con características del dominio temporal en los vectores de características. En cuanto a las señales acústicas, los bloques de actividad fueron caracterizados y clasificados a partir de la detección de breves silencios en la señal. Estos fueron detectados mediante el análisis de la energía de la señal de audio y caracterizados por la periodicidad de las pausas. Además, la adecuada combinación de etapas de procesamiento intermedio permitieron mejorar el desempeño de la estrategia básica.

Para terminar se desea destacar el nuevo enfoque utilizado para evaluar el desempeño en el reconocimiento de secuencias de actividades. Las medidas en múltiples escalas temporales brindaron información muy útil para el análisis y desarrollo de los reconocedores. Estas medidas permiten resumir y condensar los errores del reconocimiento de actividades que sólo podrían obtenerse mediante inspección visual de las etiquetas. Esto permite un completo análisis de los resultados a un nivel de detalle que no se suele realizar en otros estudios.

Como trabajos futuros se prevé continuar en los desafíos que presenta el área de ganadería de precisión para el reconocimiento de actividades. Una de las líneas posibles es estudiar la combinación de ambas modalidades (señales inerciales y acústicas), para realizar el monitoreo simultáneo de actividades alimentarias y locomotoras. Se espera además que esta combinación de modalidades pueda extenderse también a la detección de celo en vacas lecheras. Otra de las líneas posible consiste en estudiar las técnicas que permiten incorporar la dinámica de corto y largo plazo en la clasificación. Por ejemplo, se piensa evaluar la utilización de reconocedores basados en modelos ocultos de Markov o en redes recurrentes LSTM, que han resultado útiles para la clasificación de secuencias de actividades humanas.

## Referencias

- [1] R. Poppe, «A survey on vision-based human action recognition», *Image and vision computing*, vol. 28, no. 6, pp. 976–990, 2010.
- [2] J. Aggarwal and M. Ryoo, «Human activity analysis: A review», *ACM Comput. Surv.*, vol. 43, no. 3, 16:1–16:43, 2011.
- [3] O. Lara and M. Labrador, «A survey on human activity recognition using wearable sensors», *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
- [4] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, «A survey of online activity recognition using mobile phones», *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015.
- [5] Y. Guo, P. Corke, G. Poulton, T. Wark, G. Bishop-Hurley, and D. Swain, «Animal behaviour understanding using wireless sensor networks», *Proceedings of the 31st IEEE Conference on Local Computer Networks*, pp. 607–614, 2006.
- [6] D. V. Thiel and A. K. Sarkar, «Swing profiles in sport: An accelerometer analysis», *Procedia Engineering, The Engineering of Sport 10*, vol. 72, pp. 624–629, 2014.

- [7] A. Andriamandroso, J. Bindelle, B. Mercatoris, and F. Lebeau, «A review on the use of sensors to monitor cattle jaw movements and behavior when grazing», *Biotechnology, Agronomy, Society and Environment*, vol. 20, 2016.
- [8] D.-M. Tsai and C.-Y. Huang, «A motion and image analysis method for automatic detection of estrus and mating behavior in cattle», *Comput. Electron. Agric.*, vol. 104, pp. 25–31, 2014.
- [9] G. Thomas, T. B. Moeslund, and A. Hilton, «Introduction to the use of computer vision in sports», in *Computer Vision in Sports*, ser. Advances in Computer Vision and Pattern Recognition, Springer International Publishing, 2014, pp. 1–21.
- [10] P. Martiskainen, M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, «Cow behaviour pattern recognition using a three dimensional accelerometer and support vector machines», *Applied Animal Behaviour Science*, vol. 119, no. 1–2, pp. 32–38, 2009.
- [11] F. A. P. Alvarenga, I. Borges, L. Palkovič, J. Rodina, V. H. Oddy, and R. C. Dobos, «Using a three-axis accelerometer to identify and classify sheep behaviour at pasture», *Appl. Anim. Behav. Sci.*, vol. 181, pp. 91–99, 2016.
- [12] A. Bulling, U. Blanke, and B. Schiele, «A tutorial on human activity recognition using body-worn inertial sensors», *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.
- [13] T. T. Hammond, D. Springthorpe, R. E. Walsh, and T. Berg-Kirkpatrick, «Using accelerometers to remotely and automatically characterize behavior in small animals», *Journal of Experimental Biology*, pp. 1618–1624, 2016.
- [14] L. González, G. Bishop-Hurley, R. N. Handcock, and C. Crossman, «Behavioral classification of data from collars containing motion sensors in grazing cattle», *Computers and Electronics in Agriculture*, vol. 110, pp. 91–102, 2015.
- [15] C. Arcidiacono, S. Porto, M. Mancino, and G. Cascone, «Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data», *Computers and Electronics in Agriculture*, vol. 134, pp. 124–134, 2017.
- [16] V. Giovanetti, M. Decandia, G. Molle, M. Acciaro, M. Mamei, A. Cabiddu, R. Cossu, M. Serra, C. Manca, S. Rassu, *et al.*, «Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer», *Livestock Science*, vol. 196, pp. 42–48, 2017.
- [17] L. Turner, M. Udal, B. T. Larson, and S. Shearer, «Monitoring cattle behavior and pasture use with GPS and GIS», *Canadian Journal of Animal Science*, vol. 80, no. 3, pp. 405–413, 2000.

- [18] M. Li, V. Rozgić, G. Thatte, S. Lee, B. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan, «Multimodal physical activity recognition by fusing temporal and cepstral information», *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, pp. 369–380, 2010.
- [19] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, «Activity identification using body-mounted sensors—a review of classification techniques», *Physiological Measurement*, vol. 30, no. 4, R1, 2009.
- [20] L. Bao and S. Intille, «Activity recognition from user-annotated acceleration data», *Pervasive computing*, pp. 1–17, 2004.
- [21] D. Roggen, A. Calatroni, M. Rossi, T. Holleczeck, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, and A. Ferscha, «Collecting complex activity datasets in highly rich networked sensor environments», in *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*, IEEE, 2010, pp. 233–240.
- [22] P. H. Veltink, H. J. Bussmann, W. De Vries, W. L. Martens, and R. C. Van Lummel, «Detection of static and dynamic activities using uniaxial accelerometers», *Rehabilitation Engineering, IEEE Transactions on*, vol. 4, no. 4, pp. 375–385, 1996.
- [23] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depardon, «Physical activity monitoring based on accelerometry: Validation and comparison with video observation», *Medical and Biological Engineering and Computing*, vol. 37, no. 3, pp. 304–308, 1999.
- [24] H. Ichino, K. Kaji, K. Sakurada, K. Hiroi, and N. Kawaguchi, «HASC-PAC2016: large scale human pedestrian activity corpus and its baseline recognition», in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '16, USA: ACM, 2016, pp. 705–714.
- [25] M. Zhang and A. A. Sawchuk, «Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors», in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12, USA: ACM, 2012, pp. 1036–1043.
- [26] Y. Xue and L. Jin, «A naturalistic 3D acceleration-based activity dataset and benchmark evaluations», in *2010 IEEE International Conference on Systems Man and Cybernetics (SMC)*, Oct. 2010, pp. 4081–4085.
- [27] P. L. Greenwood, P. Valencia, L. Overs, D. R. Paull, and I. W. Purvis, «New ways of measuring intake, efficiency and behaviour of grazing livestock», *Anim. Produc. Sci.*, vol. 54, no. 10, pp. 1796–1804, 2014.

- [28] N. Zehner, C. Umstätter, J. J. Niederhauser, and M. Schick, «System specification and validation of a noseband pressure sensor for measurement of ruminating and eating behavior in stable-fed cows», *Computers and Electronics in Agriculture*, vol. 136, pp. 31–41, 2017.
- [29] E. A. Laca, E. D. Ungar, N. G. Seligman, M. R. Ramey, and M. W. Dement, «An integrated methodology for studying short-term grazing behaviour of cattle», *Grass Forage Sci.*, vol. 47, no. 1, pp. 81–90, 1992.
- [30] S. Navon, A. Mizrach, A. Hetzroni, and E. D. Ungar, «Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring», *Biosystems Eng.*, vol. 114, no. 4, pp. 474–483, 2013.
- [31] J. Werner, L. Leso, C. Umstatter, J. Niederhauser, E. Kennedy, A. Geoghegan, L. Shalloo, M. Schick, and B. O’Brien, «Evaluation of the RumiWatch-System for measuring grazing behaviour of cows», *J. Neurosci. Methods*, 2017.
- [32] R. J. Kilgour, «In pursuit of “normal”: A review of the behaviour of cattle at pasture», *Appl. Anim. Behav. Sci.*, vol. 138, no. 1, pp. 1–11, 2012.
- [33] J. O. Chelotti, S. R. Vanrell, D. H. Milone, S. A. Utsumi, J. R. Galli, H. L. Rufiner, and L. L. Giovanini, «A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle», *Computers and Electronics in Agriculture*, vol. 127, pp. 64–75, Sep. 2016.
- [34] D. H. Milone, J. R. Galli, C. A. Cangiano, H. L. Rufiner, and E. A. Laca, «Automatic recognition of ingestive sounds of cattle based on hidden Markov models», *Computers and Electronics in Agriculture*, vol. 87, pp. 51–55, 2012.
- [35] J. R. Galli, C. A. Cangiano, D. H. Milone, and E. A. Laca, «Acoustic monitoring of short-term ingestive behavior and intake in grazing sheep», *Livest. Sci.*, vol. 140, no. 1-3, pp. 32–41, 2011.
- [36] D. Milone, H. Rufiner, J. Galli, E. Laca, and C. Cangiano, «Computational method for segmentation and classification of ingestive sounds in sheep», *Computers and Electronics in Agriculture*, vol. 65, no. 2, pp. 228–237, 2009.
- [37] C. Goldhawk, K. Schwartzkopf-Genswein, and K. A. Beauchemin, «Validation of rumination collars for beef cattle», *J. Anim. Sci.*, vol. 91, no. 6, pp. 2858–2862, 2013.
- [38] K. Schirmann, M. A. G. von Keyserlingk, D. M. Weary, D. M. Veira, and W. Heuwieser, «Validation of a system for monitoring rumination in dairy cows», *J. Dairy Sci.*, vol. 92, no. 12, pp. 6052–6055, 2009.

- [39] Y. Meng and H.-C. Kim, «A review of accelerometer-based physical activity measurement», in *Proceedings of the International Conference on IT Convergence and Security 2011*, 120, Springer Netherlands, 2012, pp. 223–237.
- [40] C.-C. Yang and Y.-L. Hsu, «A review of accelerometry-based wearable motion detectors for physical activity monitoring», *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.
- [41] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, «Window size impact in human activity recognition», *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.
- [42] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes Ortiz, «Training computationally efficient Smartphone-Based human activity recognition models», in *Artificial Neural Networks and Machine Learning – ICANN 2013*, Springer Berlin Heidelberg, 2013, pp. 426–433.
- [43] R. San-Segundo, J. Lorenzo-Trueba, B. Martínez-González, and J. M. Pardo, «Segmenting human activities based on HMMs using smartphone inertial sensors», *Pervasive and Mobile Computing*, 2016.
- [44] A. L. H. Andriamandroso, F. Lebeau, Y. Beckers, E. Froidmont, I. Dufrasne, B. Heinesch, P. Dumortier, G. Blanchy, Y. Blaise, and J. Bindelle, «Development of an open-source algorithm based on inertial measurement units (IMU) of a smartphone to detect cattle grass intake and ruminating behaviors», *Computers and Electronics in Agriculture*, vol. 139, pp. 126–137, 2017.
- [45] M. Alsaad, J. J. Niederhauser, G. Beer, N. Zehner, G. Schuepbach-Regula, and A. Steiner, «Development and validation of a novel pedometer algorithm to quantify extended characteristics of the locomotor behavior of dairy cows», *Journal of Dairy Science*, vol. 98, no. 9, pp. 6236–6242, 2015.
- [46] W. M. Clapham, J. M. Fedders, K. Beeman, and J. P. S. Neel, «Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle», *Comput. Electron. Agric.*, vol. 76, no. 1, pp. 96–104, 2011.
- [47] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, en. John Wiley & Sons, 2012.
- [48] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: Theory and application*. Prentice Hall Englewood Cliffs, 1993, vol. 104.
- [49] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, «Direct measurement of human movement by accelerometry», *Medical Engineering & Physics*, vol. 30, no. 10, pp. 1364–1386, 2008.



- [50] M. J. Mathie, B. G. Celler, D. N. H. Lovell, and A. C. F. Coster, «Classification of basic daily movements using a triaxial accelerometer», *Medical and Biological Engineering and Computing*, vol. 42, no. 5, pp. 679–687, 2004.
- [51] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. Cardoso, «Preprocessing techniques for context recognition from accelerometer data», *Personal and Ubiquitous Computing*, vol. 14, no. 7, pp. 645–662, 2010.
- [52] J.-H. Chiang, P.-C. Yang, and H. Tu, «Pattern analysis in daily physical activity data for personal health management», *Pervasive and Mobile Computing*, vol. 13, pp. 13–25, 2014.
- [53] A. Reiss and D. Stricker, «Creating and benchmarking a new dataset for physical activity monitoring», in *Proceedings of the 5<sup>th</sup> International Conference on PErvasive Technologies Related to Assistive Environments*, USA: ACM, 2012, 40:1–40:8.
- [54] Z. He and L. Jin, «Activity recognition from acceleration data based on discrete cosine transform and svm», in *IEEE International Conference on Systems, Man and Cybernetics, 2009*, IEEE, 2009, pp. 5041–5044.
- [55] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. Lee, and J.-S. Wang, «Online classifier construction algorithm for human activity detection using a tri-axial accelerometer», *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 849–860, 2008.
- [56] S. Preece, J. Goulermas, L. Kenney, and D. Howard, «A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data», *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 871–879, 2009.
- [57] A. Godfrey, A. K. Bourke, G. M. Ólaighin, P. van de Ven, and J. Nelson, «Activity classification using a single chest mounted tri-axial accelerometer», *Medical Engineering & Physics*, vol. 33, no. 9, pp. 1127–1135, 2011.
- [58] C. A. Ronao and S.-B. Cho, «Human activity recognition with smartphone sensors using deep learning neural networks», *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.
- [59] W. Jiang and Z. Yin, «Human activity recognition using wearable sensors by deep convolutional neural networks», in *Proceedings of the 23rd ACM International Conference on Multimedia*, USA: ACM, 2015, pp. 1307–1310.
- [60] Y. Chen and Y. Xue, «A deep learning approach to human activity recognition based on single accelerometer», 2015, pp. 1488–1492.

- [61] Ó. D. Lara, A. J. Pérez, M. A. Labrador, and J. D. Posada, «Centinela: A human activity recognition system based on acceleration and vital sign data», *Pervasive and Mobile Computing*, vol. 8, no. 5, pp. 717–729, 2012.
- [62] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, «Activity classification using realistic data from wearable sensors», *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 119–128, 2006.
- [63] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, «Activity recognition of assembly tasks using body-worn microphones and accelerometers», *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 10, pp. 1553–1567, 2006.
- [64] J. Lester, T. Choudhury, and G. Borriello, «A practical approach to recognizing physical activities», in *Pervasive Computing*, ser. Lecture Notes in Computer Science 3968, Springer Berlin Heidelberg, 2006, pp. 1–16.
- [65] D. Tao, Y. Wen, and R. Hong, «Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition», *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1124–1134, 2016.
- [66] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, «Lstm: A search space odyssey», *IEEE Trans Neural Netw Learn Syst*, 2016.
- [67] M. Sokolova and G. Lapalme, «A systematic analysis of performance measures for classification tasks», *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
- [68] J. A. Ward, P. Lukowicz, and G. Tröster, «Evaluating performance in continuous context recognition using event-driven error characterisation», in *International Symposium on Location-and Context-Awareness*, Springer, 2006, pp. 239–255.
- [69] J. A. Ward, P. Lukowicz, and H. W. Gellersen, «Performance metrics for activity recognition», *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 1, 6:1–6:23, 2011.
- [70] A. Y. Yang, R. Jafari, S. S. Sastry, and R. Bajcsy, *Distributed Recognition of Human Actions Using Wearable Motion Sensor Networks*. 2009.
- [71] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán, and D. Roggen, «The opportunity challenge: A benchmark database for on-body sensor-based activity recognition», *Pattern Recognition Letters, Smart Approaches for Human Action Recognition*, vol. 34, no. 15, pp. 2033–2042, 2013.

- [72] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, «Transition-aware human activity recognition using smartphones», *Neurocomputing*, vol. 171, pp. 754–767, 2016.
- [73] A. Oppenheim and R. Schafer, «From frequency to quefrequency: A history of the cepstrum», *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.
- [74] L. R. Rabiner and R. W. Schafer, *Theory and Applications of Digital Speech Processing*, 1 edition. Upper Saddle River: Prentice Hall, 2011.



# Anexos

Referido a los tres artículos incluídos en los Anexos [A](#), [B](#) y [C](#),

- S. R. Vanrell, D. H. Milone y H. L. Rufiner, “Assessment of homomorphic analysis for human activity recognition from acceleration signals”, *IEEE Journal of Biomedical and Health Informatics*, vol. 22, n.º 4, págs. 1001-1010, 2018. doi: 10.1109/JBHI.2017.2722870
- S. R. Vanrell, J. O. Chelotti, J. Galli, H. L. Rufiner y D. H. Milone, “3D acceleration for heat detection in dairy cows”, en *Memorias de las 43º JAIIO*, Buenos Aires, Argentina, 2014, págs. 121-134.
- S. R. Vanrell, J. O. Chelotti, J. R. Galli, S. A. Utsumi, L. L. Giovanini, H. L. Rufiner y D. H. Milone, “A regularity-based algorithm for identifying grazing and rumination bouts from acoustic signals in grazing cattle”, *Computers and Electronics in Agriculture*, vol. 151, págs. 392-402, 2018. doi: 10.1016/j.compag.2018.06.021

El tesista declara haber implementado los algoritmos y llevado a cabo los experimentos descritos para obtener los resultados que allí se presentan. Estas tareas fueron realizadas bajo la guía y supervisión del director Dr. H. L. Rufiner y codirector de tesis Dr. D. H. Milone. En cuanto a la escritura de los artículos, el tesista ha sido el autor principal, guiado por los comentarios, sugerencias y revisiones del director, codirector de tesis y otros coautores que en cada artículo se indican.

Los abajo firmantes avalan esta declaración.

---

Dr. H. L. Rufiner

Director

---

Dr. D. H. Milone

Co-director

En el marco del proyecto de ganadería de precisión se presentaron 3 solicitudes de patentes donde el tesista es uno de los inventores.

- Patente nacional. Inventores: J. R. Galli, L. L. Giovanini, D. H. Milone, H. L. Rufiner, J. O. Chelotti y S. R. Vanrell, “Proceso para detectar celo en un animal rumiante y dispositivo”. Titulares: UNL, CONICET, UNR. Expediente: P20160103881. Fecha de aplicación: 2016-12-16
- Patente internacional. Inventores: L.L. Giovanini, H.L. Rufiner, D.H. Milone, J.R. Galli, J.O. Chelotti y S.R. Vanrell, “A method and a device for detecting, classifying and quantifying in real time the feeding activity of ruminant animals”. Titulares: UNL, CONICET, UNR. Expediente: PCT/IB2016/057627. Fecha de aplicación: 2016-12-14
- Patente nacional. Inventores: JL.L. Giovanini, H.L. Rufiner, D.H. Milone, J.R. Galli, J.O. Chelotti y S.R. Vanrell, “Proceso para detectar celo en un animal rumiante y dispositivo”. Titulares: UNL, CONICET, UNR. Expediente: P20150104082. Fecha de aplicación: 2015-12-15

El tesista es integrante de dicho proyecto y ha sido partícipe de la discusión de las ideas allí incluídas así como de los experimentos relacionados. Por ello ha colaborado en la revisión y escritura de las mismas. Relacionado a los problemas enfrentados en esta tesis, se incluye en el Anexo D una copia de la descripción técnica de la patente nacional P20160103881. En esta patente, los principales aportes del tesista se dieron en las estrategias propuestas para el monitoreo del comportamiento individual.

Los abajo firmantes avalan esta declaración.

---

Dr. H. L. Rufiner  
Director

---

Dr. D. H. Milone  
Co-director

## Anexo A

# Assessment of homomorphic analysis for human activity recognition from acceleration signals

El artículo ha sido publicado en la revista IEEE Journal of Biomedical and Health Informatics.

S. R. Vanrell, D. H. Milone y H. L. Rufiner, “Assessment of homomorphic analysis for human activity recognition from acceleration signals”, *IEEE Journal of Biomedical and Health Informatics*, vol. 22, n.º 4, págs. 1001-1010, 2018.

doi: 10.1109/JBHI.2017.2722870





# Assessment of Homomorphic Analysis for Human Activity Recognition from Acceleration Signals

Sebastián R. Vanrell<sup>1</sup>, Diego H. Milone<sup>1</sup>, Hugo L. Rufiner<sup>1,2</sup>

<sup>1</sup> Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL/CONICET, Argentina.

<sup>2</sup> Laboratorio de Cibernética, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina

Received: April 19, 2017; Accepted: June 28, 2017; Published: July 3, 2017

**Abstract.** Unobtrusive activity monitoring can provide valuable information for medical and sports applications. In recent years, human activity recognition has moved to wearable sensors to deal with unconstrained scenarios. Accelerometers are the preferred sensors due to their simplicity and availability. Previous studies have examined several classic techniques for extracting features from acceleration signals, including time-domain, time-frequency, frequency-domain, and other heuristic features. Spectral and temporal features are the preferred ones and they are generally computed from acceleration components, leaving the acceleration magnitude potential unexplored. In this study, a new type of feature extraction stage, based on homomorphic analysis, is proposed in order to exploit discriminative activity information present in acceleration signals. Homomorphic analysis can isolate the information about whole body dynamics and translate it into a compact representation, called cepstral coefficients. Experiments have explored several configurations of the proposed features, including size of representation, signals to be used, and fusion with other features. Cepstral features computed from acceleration magnitude obtained one of the highest recognition rates. In addition, a beneficial contribution was found when time-domain and moving pace information was included in the feature vector. Overall, the proposed system achieved a recognition rate of 91.21% on the publicly available SCUT-NAA dataset. To the best of our knowledge, this is the highest recognition rate on this dataset.

**Index terms**— accelerometer, cepstrum, human activity recognition, signal processing.

# 1. Introduction

Research interest in human activity recognition (HAR) has increased in the last two decades. Nowadays, HAR applications spread over physical monitoring [1]–[4], health [1], [5]–[7], entertainment [5], [8], [9], sports [10], security [5], [9], [11], and industry [9], [12]. For instance, activities of patients can be reliably tracked, which can help a physician to counsel behaviors for physical activity and healthy lifestyle [3], [6]. Historically, research in computer vision has been in the vanguard of HAR [5], [9]. However, efforts to recognize activities in unconstrained daily life settings caused a shift toward the use of wearable sensors. Accelerometers, gyroscopes, and compasses are the sensors typically used, either individually or combined. In this work, we will focus on HAR systems that consider accelerations as input signals.

The first studies on activity recognition using accelerometers were published in the middle and late 1990s, with applications in medical assessment [13]. They used unidirectional accelerometers on two or three locations of the body, and applied a set of simple rules to distinguish between dynamic and static activities. At present, HAR systems are based on triaxial accelerometers and they aim to recognize a broader set of activities. These systems rely on pattern recognition, in which the fundamental stages include signal pre-processing, feature extraction, and activity classification. Several proposals for each of these stages have been made to improve said systems [1], [2], [8], [14]–[16].

Ideally, a feature extraction stage should be able to extract all the discriminative information in a compact representation. Discriminative information helps to distinguish activities from one another. For example, the periodicity combined with a spectral description of movements are good candidates. A compact representation helps to keep the recognizer as simpler as possible and avoids high computational costs. Previous studies have explored a wide range of techniques for extracting features from acceleration signals. However, features related to activities are not discriminative enough and others are not compact. In this study a compact representation is proposed and its potential is fully assessed.

Feature extraction techniques explored in previous studies can be grouped by the type of their outputs as heuristic, time-domain, time-frequency, or frequency-domain features [14]. Heuristic features are derived from and characterized by an intuitive understanding of how an activity produces changes in the signal. For example, a static posture such as standing or lying can be recognized by the direction of the recorded gravity [17]. Other examples are signal magnitude area, mean rectified value, peak-to-peak acceleration, and root mean square, which have been related to the intensity of an activity [18].

In contrast to heuristic features, time-domain features are not directly related to specific aspects of individual movements or postures. Instead, they are computed from a windowed signal and they are typically statistical measures. Com-

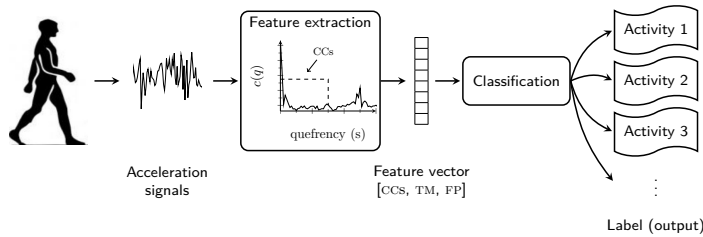


Figure 1: General diagram of the proposed recognizer with its two main stages: feature extraction and classification. CCs: cepstral coefficients; TM: time-domain measures; FP: fundamental period.

mon examples include mean, median, variance, skewness, kurtosis, interquartile range, and correlation between accelerometer axes [14], [19], [20]. Recently, a two-directional feature for incremental learning showed promising results [21]. Both time-domain and heuristic features are the simplest to compute, and they were useful for distinguishing between static and dynamic activities. However, they were barely successful in discriminating between dynamic activities [14], [22].

Time-frequency features, such as wavelets, were effective for detecting transitions between activities, but they were outperformed by other features in the activity classification task [23]. By contrast, frequency-domain features are usually the preferred choice in HAR [19], [22], [24]–[27]. These features are obtained using the fast Fourier transform (FFT) or the discrete cosine transform. Basis coefficients may be directly adopted as features or an additional method may be used to characterize the spectral distribution, such as subsets of coefficients, or filter banks. Several studies have reported their best results using FFT coefficients as features, either individually [19], [22], [25], [26] or fused with time-domain features [20], [24], [28]. In these studies, features are usually extracted from each of the acceleration-component signals, leaving the capabilities of acceleration-magnitude signal unexplored, which have processing and storage advantages.

The feature extraction stage proposed in this work uses a well-known technique: the cepstral analysis. It is based on the theory of homomorphic analysis and has been successfully used for characterizing seismic echoes, deblurring images, and recognizing speech [29], but it has not been considered on HAR, except for two studies [30], [31]. However, these studies have neither examined the compliance of hypothesis to apply such analysis nor exploited all the information that cepstral representation can capture. Li *et al.* [31] claimed that signals have quasi-periodic characteristics although the analysis was applied on narrow windows, which cannot encompass that periodicity. In [30], a mel scale was used to map frequency bands in a nonlinear fashion. This mel scale was designed for speech analysis and recognition based on the perceptual scale in the human cochlea [29]. Although

this makes no sense for HAR, a high resolution at low frequencies could be useful. In a related study [32], the features proposed in [30] were also used for user identification. For this task, adapted perceptual linear prediction coefficients were the best alternative.

In the present study, the proposed representation explores the capabilities of cepstral coefficients as is (i.e., without using intermediate filter banks or dimension reduction techniques). In addition, it is the first time that body dynamics and activity periodicity are explained in terms of cepstral features. The advantage of the proposed cepstral representation is that it compresses the spectral information of body dynamics in few coefficients while retaining its discriminative power. In the experiments, features extracted from magnitude and components of acceleration were contrasted based on their discriminative power and an exhaustive analysis on the number of required coefficients was performed. Furthermore, we assessed the performance improvements when including time-domain measures and the fundamental period of the signal. To complete the system, a support vector machine was chosen for the classification stage, because it showed better performance compared to other well-known classifiers, such as multilayer perceptron, random forest, and naive Bayes. In addition, the selection of support vector machine allows a fair comparison with previous studies.

The remainder of this paper is organized as follows. Section 2 describes the proposed HAR system. Section 3 presents the experimental setup. Results and discussion are given in Section 4. Finally, conclusions are drawn in Section 5.

## 2. Proposed recognition system

As most automatic recognition systems, the proposed recognizer can be seen as a pipeline with two main stages: feature extraction and classification. An schematic diagram of the system based on homomorphic analysis is shown in Figure 1. The first stage will be explained in detail from Section 2.1 to 2.3. The second stage will be covered in Section 2.4.

### 2.1. Acceleration signals

In this study, a three-dimensional accelerometer is used for capturing information about movements. The captured signals are the acceleration components  $a_x(t)$ ,  $a_y(t)$ , and  $a_z(t)$ . In addition, the acceleration magnitude  $|\mathbf{a}(t)|$  is computed from components as another input signal. Components are recorded in directions relative to the orientation of the device, thus being altered by gravity when the device rotates. On the contrary, magnitude is unaltered by the orientation of the

device because gravity is a constant offset. These signals are processed as described below for a generic signal called  $a(t)$ .

In what follows the acceleration signal  $a(t)$  is considered as the output of a linear convolutive system. The excitation  $m(t)$  is associated to the moving pace, which is originated by muscle activity and external body interactions. Body dynamics is modeled in the impulse response  $h(t)$ , which will depend on the activity that is being performed. Since the goal is to recover information about said activity, the analysis should be able to isolate  $h(t)$ . However, excitation and impulse response are unknown; therefore recovering  $h(t)$  is not straightforward and a blind deconvolution is required. We propose using a homomorphic analysis to carry out the recovering task.

## 2.2. Cepstral coefficients

Homomorphic analysis was developed as a general method for separating signals that have been non-additively combined [29]. Homomorphic processing involves converting this mixture into a linear combination, in which the analysis techniques are well understood. For instance, if two signals are convolved in time-domain as

$$a(t) = m(t) * h(t),$$

their Fourier transforms will be multiplied in frequency-domain as

$$A(f) = M(f) \cdot H(f),$$

and an appropriately defined logarithm will produce the sum

$$\log |A(f)| = \log |M(f)| + \log |H(f)|.$$

At this point,  $\log |M(f)|$  and  $\log |H(f)|$  are additively combined. Taking the inverse Fourier transform of  $\log |A(f)|$ , a new time-representation,

$$c(q) = \mathcal{F}^{-1} \{ \log |A(f)| \},$$

can be obtained. This is known as the cepstrum and its domain is the quefrequency. At the beginning of  $c(q)$ , the low-rate variations of  $\log |A(f)|$  are codified. The remaining information of  $c(q)$  codifies the high-rate variations of  $\log |A(f)|$ . Generally, low- and high-rate variation components are separated, and they can be linked to the impulse response and the excitation, respectively. Thus, by taking the first coefficients of  $c(q)$ , the homomorphic analysis can recover the desired information about the underlying system. In the quefrequency domain, this operation is known as liftering.

In the time domain, acceleration signals exhibit a quasi-periodic behavior, as exemplified in Figure 2.a. For this example, the period is  $\sim 0.5$  s, which is the expected moving pace for a daily human activity. This signal gives some insight into the excitation signal  $m(t)$ , but the impulse response  $h(t)$  remains hidden. In the frequency domain, this period can be seen as a peak of fundamental frequency at  $\sim 2$  Hz and its corresponding harmonics (Figure 2.b). These are the major contributions, but they mask the remaining information about the impulse response of the system. Figure 2.c shows the signal obtained by applying the logarithm to  $|A(f)|$ . Afterwards, the inverse Fourier transform is applied and the cepstrum,  $c(q)$ , is obtained (Figure 2.d). Peaks related to excitation are still present at  $\sim 0.5$  s and  $\sim 1.0$  s but no longer overlapped with the cepstrum of impulse response.

For example, if  $a(t)$  is captured on the hip of a subject performing an activity, the first coefficients of  $c(q)$  codify the dynamics of hip movements. The hip was reported as the best position to represent the whole body, thus the coefficients at the left of the first peak codify the global dynamics of the body [2], regardless of the activity pace. Besides, the separation between peaks in  $c(q)$  represents the pace at which the activity is performed.

As explained above, the beginning of  $c(q)$  contains the low-rate variations of  $\log |A(f)|$ , which is the isolated information about  $h(t)$ . To verify this, a smooth representation of  $\log |A(f)|$  can be recovered by applying the Fourier transform only to the first coefficients of  $c(q)$ , i.e., the transform is applied to a lifted  $c(q)$ . The lifters used in the recovering are indicated by dashed rectangles in Figure 2.d and the smoothed versions of  $\log |A(f)|$  are plotted in Figure 2.e using the respective colors. These curves are estimations of  $\log |H(f)|$ . It is clear that the number of coefficients defines the degree of details retained in the recovered information about the impulse response of the body dynamics.

Computation of cepstral coefficients requires the application of a Fourier transform, a logarithm and an inverse Fourier transform. These transformations require  $O(n \log(n))$ ,  $O(n)$ , and  $O(n \log(n))$ , respectively, where  $n$  is the length of the signal analyzed. Thus, cepstral features computation is of the same order as any other FFT-based feature, which requires  $O(n \log(n))$ .

### 2.3. Feature vector composition

In this study, the analysis of a signal  $a(t)$  may lead to three kinds of features: cepstral coefficients, time-domain measures, and fundamental period. Cepstral coefficients (CCs) refer to the first coefficients of  $c(q)$ , up to a predefined quefrency. Time-domain measures (TM) are five values computed directly from  $a(t)$ : standard deviation, energy, maximum amplitude, minimum amplitude, and peak-to-peak amplitude. The fundamental period (FP) contains a value that represents the moving pace.

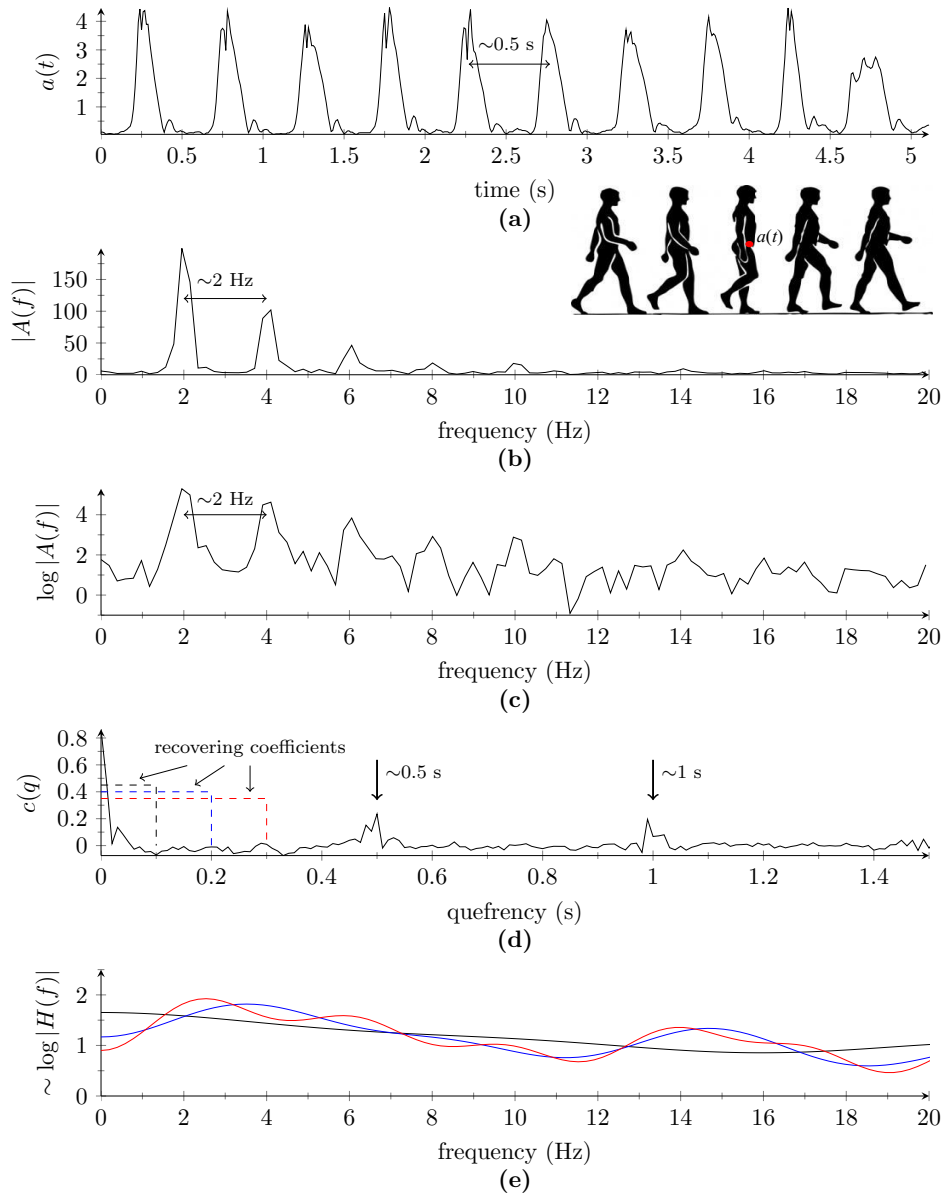


Figure 2: Computation of cepstral coefficients. (a) Quasi-periodic acceleration signal  $a(t)$ ; (b) spectrum of  $a(t)$ ; (c) logarithm of the spectrum; (d) cepstral coefficients; (e) recovered approximations of  $\log |H(f)|$ .

Table 1: Composition of feature vectors when cepstral coefficients (CC) are obtained from  $a_x(t)$ ,  $a_y(t)$ , and  $a_z(t)$ . FP: fundamental period. TM: time-domain measures.

Feature	Signal	Composition of feature vectors															
		F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>	F <sub>7</sub>	F <sub>8</sub>	F <sub>9</sub>	F <sub>10</sub>	F <sub>11</sub>	F <sub>12</sub>	F <sub>13</sub>	F <sub>14</sub>	F <sub>15</sub>	F <sub>16</sub>
CC	$a_*$	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
TM	$ \mathbf{a} $		•		•		•		•		•		•		•		•
FP	$ \mathbf{a} $			•	•			•	•			•	•			•	•
TM	$a_*$					•	•	•	•				•	•	•	•	•
FP	$a_*$									•	•	•	•	•	•	•	•

The example in the previous section shows that FP is revealed as one or more peaks in  $c(q)$  (Figure 2.d). However, there are activities that present an unclear or absent peak. Therefore, we chose to extract FP with the unbiased autocorrelation of  $a(t)$  [33]. To be considered as a periodical activity, the amplitude of autocorrelation must exceed a given threshold. In such a case, the lag corresponding to the first local maximum is assigned to FP. On the contrary, FP will be set to zero, which means that no periodicity was detected.

As  $a(t)$  can be  $a_x(t)$ ,  $a_y(t)$ ,  $a_z(t)$ , or  $|\mathbf{a}(t)|$ , each of these signals can generate its own set of features. Thus, different combinations could be selected to define the final feature vector. Tables 1 and 2 summarize the combinations evaluated in this study. Vectors composed by cepstral coefficients of  $a_x(t)$ ,  $a_y(t)$ , and  $a_z(t)$  are given in Table 1. Vectors composed by cepstral coefficients of  $|\mathbf{a}(t)|$  are presented in Table 2. These tables present the names assigned to each composition, which will be used in Section 4. Combinations were selected to perform a comparative study which can assess the partial contribution of each feature. First, the study will evaluate the influence of the number of CCs, without additional features (F<sub>1</sub> in Table 1 and F<sub>17</sub> in Table 2). Then, the contributions of temporal and pace information will be considered in F<sub>2</sub>, F<sub>3</sub> and F<sub>4</sub> (Table 1), and F<sub>18</sub>, F<sub>19</sub> and F<sub>20</sub> (Table 2). The study will also consider the influence of the signal which gave origin to a feature. For instance, F<sub>2</sub> and F<sub>5</sub> are comprised by the same CCs but differ on the source signals to compute TM ( $|\mathbf{a}|$  or  $a_*$ ). Finally, combined effects will be studied by considering all the combinations of these variants.

The above description assumes that there is only one three-dimensional acceleration signal of short duration. However, each of these signals may be of long and variable duration. Thus, a collection of acceleration signals  $\mathbf{a}^1(t), \mathbf{a}^2(t), \dots$  is obtained from a raw acceleration data by a fixed-size sliding window, with overlap between consecutive windows (top of Figure 3). Therefore, the feature extraction process generates a sequence of feature vectors  $X = \{\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \dots, \mathbf{x}^L\}$ , where  $L$



Table 2: Composition of feature vectors when cepstral coefficients (CC) are obtained from  $|\mathbf{a}(t)|$ . FP: fundamental period. TM: time-domain measures.

Feature	Signal	Composition of feature vectors															
		F <sub>17</sub>	F <sub>18</sub>	F <sub>19</sub>	F <sub>20</sub>	F <sub>21</sub>	F <sub>22</sub>	F <sub>23</sub>	F <sub>24</sub>	F <sub>25</sub>	F <sub>26</sub>	F <sub>27</sub>	F <sub>28</sub>	F <sub>29</sub>	F <sub>30</sub>	F <sub>31</sub>	F <sub>32</sub>
CC	$ \mathbf{a} $	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
TM	$ \mathbf{a} $		•		•		•		•		•		•		•		•
FP	$ \mathbf{a} $			•	•			•	•			•	•			•	•
TM	$a_*$					•	•	•	•				•	•	•	•	•
FP	$a_*$									•	•	•	•	•	•	•	•

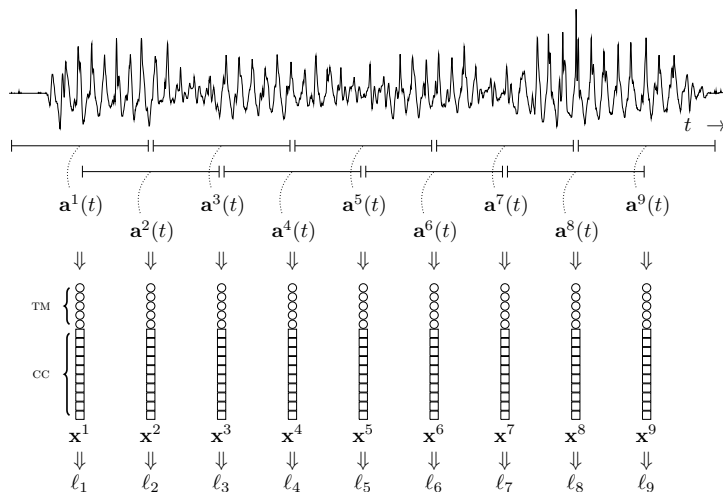


Figure 3: Sequence of feature vectors generated from a raw acceleration signal. In this example, feature vectors are composed by time-domain measures (TM) and ten cepstral coefficients (CC) extracted from the acceleration magnitude.

is the number of windows (bottom of Figure 3).

## 2.4. Classification

In the classification stage, a well-known static classifier is used: the support vector machine (SVM) [34]. This classifier has proven to be one of the best machine-learning techniques for binary classification problems [35]. First, input features are non-linearly mapped to a very high-dimension feature space. The mapping is done using a kernel function. Then, a linear decision surface is built in this new feature space. That surface is a hyperplane and it is located so that the maximum margin of separation is achieved between the two classes (see Hastie *et al.* [36] for a detailed explanation). In this study, a radial basis function  $K(\mathbf{x}, \mathbf{y}) = e^{-\gamma|\mathbf{x}-\mathbf{y}|^2}$

was chosen as a kernel, where  $\mathbf{x}$  and  $\mathbf{y}$  are feature vectors. Other kernel choices were discarded in preliminary experiments. Also, a soft margin penalty for misclassifications was considered. Penalty coefficient  $C$  and parameter  $\gamma$  were optimized with a logarithmic grid search on the training dataset. In the proposed system, the number of classes is equal to the number of activities to recognize. For this multiclass classification task, the one-against-one approach was followed [37], [38].

An SVM requires a number of operations during training that is different from classification time. During training, complexity of an SVM depends greatly on the number of vectors in the dataset, rather than the dimension of those vectors. In addition, SVM can perform very well with relatively large feature vectors. During classification, the number of required operations is linear with the number of support vectors and the dimension of feature vectors.

Feature selection may be included prior to classification and there is a vast diversity of techniques proposed in the literature for this task. For instance, there are techniques that can be applied to any kind of classifiers [39] and others that are embedded and specifically designed for SVM [40]. However, there is no clear definition on which technique is suitable in advance. In general, it depends on chosen classifier and dataset characteristics, such as number of examples and inter-relationship among features. A thorough evaluation of selection methods with cepstral representation falls outside the scope and extension of the present study.

As was previously described, a raw acceleration data corresponding to a single activity is translated into a sequence of feature vectors  $X = \{\mathbf{x}^1, \mathbf{x}^2, \dots\}$ . SVM classifies each vector  $\mathbf{x}^i$  independently from one another and assigns an activity label  $\ell_i$  to each one. Then, the label count is computed by activity, and the label that corresponds to the maximum count is assigned to the input acceleration data.

### 3. Experimental setup

The recognition experiments were carried out on the publicly available SCUT-NAA dataset [22], which contains 1278 records from 44 subjects collected in naturalistic settings. During the recording sessions, subjects were asked to perform ten activities, one time each, thereby being a dataset with balanced classes. Each recorded signal represented a single activity (i.e., the provided signals were already segmented by activity), with sample frequency fixed at 100 Hz. Subjects wore a triaxial accelerometer on their waist belts, trouser pockets, and shirt pockets, alternately. Previous studies have demonstrated that the waist is the best location for a single sensor because it can better represent the major human motion [2]. Thus, only records captured on waist belt position will be used in the following experiments. Features extracted considering this position will contain information of the whole body dynamics.

Features were extracted from raw acceleration signals using a window size of 5.12 s. The overlapping between consecutive windows was 50%. Signals were sampled at 100 Hz, hence, windows size was of 512 samples, with 256 samples overlapping between consecutive windows. Each segment was smoothed with a Hamming window. Several cepstral lengths were tested: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 s (in quefrency), which correspond to  $n_{cc} = \{10, 20, 30, \dots, 90\}$  coefficients, respectively. Regardless of final cepstrum length, time-domain signals were transformed with a 512-points FFT. A previous study [31] performed cepstral analysis with narrow windows (<1 s) that cannot capture a single period of an activity. Also, the resolution of cepstral representation was not explored. Authors in [30] tested different length of their speech-adapted representation but the performance of isolated acceleration features was not studied.

In this study, leave-one-subject-out cross-validation was used to conduct the experiments, which allows a fair comparison with previous results on the same dataset. In the first fold of this scheme, signals from subject 1 were taken for testing, while the remaining signals (from the other subjects) were used for selecting the best parameters of the classifier and for training the models. The following folds switch the test signals to another subject until all subjects are considered in the test. As in previous studies [19], [22], activity recognition rate will be used as a performance measure, which is simply defined as the number of correctly classified signals over the total number of classified signals. Also, the standard recall ( $t_p/[t_p + f_n]$ ) and precision ( $t_p/[t_p + f_p]$ ) are reported on confusion matrices, where  $t_p$  are true positives,  $f_p$  false positives, and  $f_n$  false negatives counts for class  $c$ . In addition to standard measures, the relative error reduction will be used to compare the performance of different systems. This measure is useful when the recognition rates are close to 100%. For two recognizers with absolute errors  $\epsilon_A$  and  $\epsilon_B$ , where  $A$  is the reference system, the relative error reduction is  $\delta = (\epsilon_B - \epsilon_A)/\epsilon_A$ .

## 4. Results and discussion

The following subsections present a detailed analysis of the results obtained. First, feature vector compositions are compared based on their discriminative power. Second, these compositions are distinguished from other approaches based on cepstral analysis. Third, the proposed system is compared with state-of-the-art systems. Fourth, a related recognition task that joins similar activities is formulated. For this modified task, the proposed system and some adaptations are evaluated and discussed.

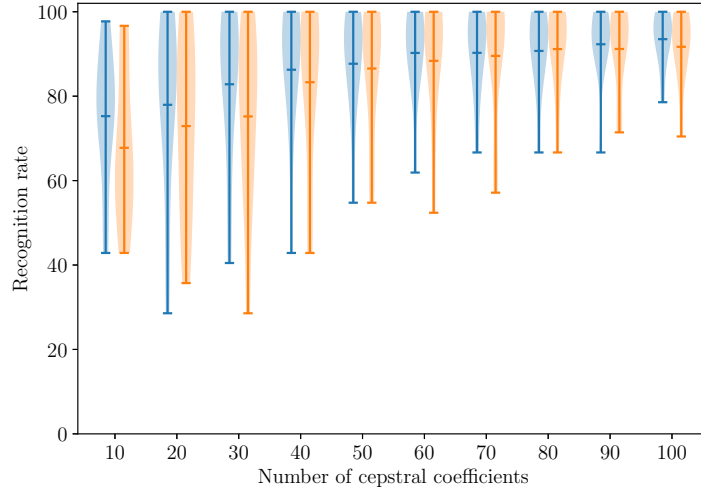


Figure 4: Violin plots of the recognition rate per activity versus the number of CCs. Feature vector composition: (blue) CCs of the acceleration components (F<sub>1</sub>), (orange) CCs of the acceleration magnitude (F<sub>17</sub>).

#### 4.1. Comparative analysis of feature vector compositions

Feature vectors can comprise features extracted from  $|\mathbf{a}(t)|$ ,  $a_x(t)$ ,  $a_y(t)$ , and  $a_z(t)$ . A comparative study was conducted to select the most suitable composition. This included the influence of the number of CCs, and the contributions of temporal and pace information. Both points will be addressed in the following paragraphs.

The influence of the number of CCs on the recognition rate per activity is shown in the violin plots of Figure 4 for F<sub>1</sub> and F<sub>17</sub>. The composition F<sub>1</sub> corresponds to the CCs obtained from the acceleration components and F<sub>17</sub> corresponds to the CCs obtained from the acceleration magnitude (Tables 1 and 2). As more coefficients are selected (from 10 to 90), the mean of recognition rates improves and the distributions get narrower. Particularly, improvements are obtained with up to 50 coefficients. A possible explanation is that this resolution of the cepstral representation can capture most of the discriminative information in the impulse response. Also, discriminative cues can be associated with the first periodicity peak that typically appears between the 35th and the 50th cepstral coefficient (i.e., 0.35 and 0.50 s at 100 Hz sampling frequency). Using more coefficients showed no significant improvements. Hence, up to 70 CCs will be considered in the following experiments.

The possible compositions were expanded when other features are appended to CCs. The results for all of the proposed compositions when CCs are extracted from  $a_x(t)$ ,  $a_y(t)$ , and  $a_z(t)$  are compiled in Table 3. The number of CCs is specified in

Table 3: Activity recognition performance when CCs are obtained from  $a_x(t)$ ,  $a_y(t)$ , and  $a_z(t)$ .

Features (size)	$n_{cc}$ : Number of cepstral coefficients						
	10	20	30	40	50	60	70
F <sub>1</sub> (0+3 $n_{cc}$ )	74.82	77.43	82.42	85.99	87.41	<b>90.02</b>	90.02
F <sub>2</sub> (5+3 $n_{cc}$ )	78.86	80.76	87.17	86.94	87.65	89.07	<b>89.31</b>
F <sub>3</sub> (1+3 $n_{cc}$ )	79.57	82.66	85.51	88.12	88.60	90.74	<b>90.97</b>
F <sub>4</sub> (6+3 $n_{cc}$ )	83.85	84.09	88.36	87.89	88.84	<b>90.02</b>	90.02
F <sub>5</sub> (15+3 $n_{cc}$ )	80.05	84.32	85.75	87.41	87.89	<b>89.31</b>	89.31
F <sub>6</sub> (20+3 $n_{cc}$ )	80.05	83.61	85.75	87.41	87.89	87.41	<b>88.60</b>
F <sub>7</sub> (16+3 $n_{cc}$ )	83.37	86.94	88.12	89.31	88.60	89.07	<b>89.79</b>
F <sub>8</sub> (21+3 $n_{cc}$ )	83.85	86.94	88.36	88.60	88.60	88.12	<b>90.12</b>
F <sub>9</sub> (3+3 $n_{cc}$ )	79.81	84.56	84.32	86.22	88.84	89.79	<b>90.26</b>
F <sub>10</sub> (8+3 $n_{cc}$ )	81.47	84.80	86.46	87.41	87.65	<b>89.07</b>	89.07
F <sub>11</sub> (4+3 $n_{cc}$ )	80.29	84.32	86.46	87.41	89.79	<b>90.26</b>	90.26
F <sub>12</sub> (9+3 $n_{cc}$ )	81.00	85.75	87.89	88.84	88.12	89.31	<b>89.55</b>
F <sub>13</sub> (18+3 $n_{cc}$ )	83.61	85.99	86.70	86.70	<b>88.84</b>	88.60	88.84
F <sub>14</sub> (23+3 $n_{cc}$ )	83.14	85.99	87.17	87.65	88.36	88.60	<b>88.84</b>
F <sub>15</sub> (19+3 $n_{cc}$ )	84.32	86.94	88.36	87.41	88.84	88.84	<b>89.79</b>
F <sub>16</sub> (24+3 $n_{cc}$ )	82.90	87.41	88.36	88.12	88.60	89.07	<b>89.79</b>

the header of the table. The results in each row correspond to an exclusive composition that only differs in the number of CCs. The improvements in recognition rates are clearly higher when TM or FP are appended to a small number of CCs, because they compensate the lack of discriminative power of low-resolution cepstral representations. However, when more CCs are used, the cepstral representation by itself achieves high recognition rates and the contribution of other features is reduced. For example, if 10 CCs are used (F<sub>1</sub>), the recognition rate is 74.82%, and the rate improves to 83.85% when TM and FP (F<sub>4</sub>) are added. However, if 70 CCs are selected (F<sub>1</sub>), the recognition rate is 90.02%, and no improvement is observed when TM and FP (F<sub>4</sub>) are added. In the absence of TM, FP helps to distinguish between activities, achieving a high rate of 90.97% for 70 CCs (F<sub>19</sub>).

Table 4 presents the results for CCs extracted from  $|\mathbf{a}(t)|$  (F<sub>17</sub>-F<sub>32</sub>). Clear improvements are achieved by appending TM or FP obtained from  $|\mathbf{a}(t)|$  to CCs (compare the results for F<sub>18</sub>-F<sub>20</sub> with F<sub>17</sub>). Time-domain measures help to avoid confusion between activities that are clearly different in their waveform. For example, jumping involves a high acceleration that must oppose to gravity. Thus, the recorded signal  $|\mathbf{a}(t)|$  will exhibit a great amplitude and energy that can be easily distinguished from the other activities. However, adding TM slightly improves the discriminability of similar activities. For example, activities such as walking and

Table 4: Activity recognition performance when CCs are obtained from  $|\mathbf{a}(t)|$ .

Features (size)	$n_{cc}$ : Number of cepstral coefficients						
	10	20	30	40	50	60	70
F <sub>17</sub> (0+ $n_{cc}$ )	66.98	72.21	74.58	82.90	86.22	88.12	<b>89.31</b>
F <sub>18</sub> (5+ $n_{cc}$ )	75.53	83.37	85.99	85.51	88.12	87.89	<b>88.36</b>
F <sub>19</sub> (1+ $n_{cc}$ )	76.72	80.52	83.85	84.32	87.65	88.60	<b>88.84</b>
F <sub>20</sub> (6+ $n_{cc}$ )	82.66	85.04	87.17	87.65	88.84	89.07	<b>90.02</b>
F <sub>21</sub> (15+ $n_{cc}$ )	80.76	84.56	84.56	85.51	86.70	88.84	<b>89.31</b>
F <sub>22</sub> (20+ $n_{cc}$ )	82.66	85.75	84.80	87.65	88.84	89.31	<b>90.97</b>
F <sub>23</sub> (16+ $n_{cc}$ )	83.61	88.84	86.70	87.65	88.12	<b>90.02</b>	88.84
F <sub>24</sub> (21+ $n_{cc}$ )	84.09	87.41	87.41	88.36	90.02	90.50	<b>91.21</b>
F <sub>25</sub> (3+ $n_{cc}$ )	76.01	82.19	83.61	85.04	88.36	<b>89.79</b>	88.84
F <sub>26</sub> (8+ $n_{cc}$ )	80.05	85.51	86.94	87.65	<b>89.07</b>	89.07	89.07
F <sub>27</sub> (4+ $n_{cc}$ )	75.77	82.66	84.56	85.99	88.84	<b>89.31</b>	89.31
F <sub>28</sub> (9+ $n_{cc}$ )	81.95	86.46	88.84	88.60	<b>90.50</b>	90.50	90.02
F <sub>29</sub> (18+ $n_{cc}$ )	85.27	86.70	88.36	85.99	87.17	<b>89.31</b>	88.84
F <sub>30</sub> (23+ $n_{cc}$ )	84.56	87.89	88.36	89.07	<b>89.79</b>	88.84	88.12
F <sub>31</sub> (19+ $n_{cc}$ )	85.27	87.89	87.65	87.89	87.17	<b>89.07</b>	88.36
F <sub>32</sub> (24+ $n_{cc}$ )	83.61	87.65	88.36	88.60	<b>90.50</b>	89.31	89.31

walking quickly are expected to be very similar in their temporal and cepstral representations. Fundamental period can be of help in this situation, when repetitive actions are performed at different paces. As a result, it is observed that, regardless of the number of CCs, the recognition rates are higher when TM and FP (F<sub>20</sub>) are combined than when they are used individually.

As presented in Table 3, the improvements in recognition rates are greater if TM and FP are appended to low-resolution cepstral representations rather than high-resolution representations. It is evident that discriminative information of both features is overlapped, since the improvement is lower when they are combined (F<sub>20</sub>) than when they are used individually (F<sub>18</sub> or F<sub>19</sub>). In addition, a recognition rate of 90.02% is an interesting result for a composition (F<sub>20</sub>) that uses features exclusively obtained from  $|\mathbf{a}(t)|$ , because this signal is independent of the orientation of the recording device. Moreover, this result is comparable to the ones presented in Table 3, although it was obtained with a smaller feature vector (corresponding sizes are specified in Supplementary Material).

Additional information can be included in feature vectors by appending TM and FP from acceleration components (F<sub>21</sub>-F<sub>32</sub>). For example, activities such as walking and walking upstairs involve only one direction, thereby affecting only one particular component of the acceleration. Walking primarily modifies the acceleration in the axis related to forward movement, while walking upstairs affects axes

Table 5: Performance comparison of different classifiers for feature vector  $F_{24}$  with 70 CCs.

Activity	SVM	NaiveBayes	Random Forest	MLP
Step walking (s)	<b>97.73</b>	13.64	90.91	95.45
Jumping (j)	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
Bicycling (b)	<b>100.00</b>	36.67	<b>100.00</b>	<b>100.00</b>
Walking (w)	<b>75.00</b>	0.00	65.91	<b>75.00</b>
Walking backward (wb)	<b>90.70</b>	83.72	69.77	<b>90.70</b>
Walking quickly (wq)	<b>76.19</b>	2.38	66.67	73.81
Running (r)	<b>100.00</b>	93.18	<b>100.00</b>	<b>100.00</b>
Relaxing (re)	97.73	<b>100.00</b>	97.73	97.73
Downstairs (d)	<b>86.36</b>	72.73	81.82	81.82
Upstairs (u)	<b>90.70</b>	81.40	81.40	79.07
Overall	<b>91.21</b>	59.14	85.04	89.07

related to vertical and forward movements. Thus, temporal information of acceleration components could reveal unseen differences between activities. This may explain why higher recognition rates are achieved with  $F_{22}$ ,  $F_{24}$ , and  $F_{28}$  compared to  $F_{20}$ . Although compositions that use TM and FP obtained from components ( $F_{29}$ - $F_{32}$  in Table 4) achieve good recognition rates, they are not as good as the ones mentioned above. Specifically, said compositions show their best results for 50 and 60 CCs. This may be explained by the combined contributions of TM and FP from components, which compensate the loss from the use of a slightly coarse cepstral representation.

The best recognition rate (91.21%) was achieved with the feature vector  $F_{24}$ . Other well-known classifiers such as Naive Bayes, Random Forest, and Multilayer Perceptron were tested but none of them showed better results than SVM. Table 5 compare activity recall and overall recognition rate for  $F_{24}$  composition and these classifiers. This composition combines the best of all the proposed features. It has the best cepstral representation (70 CCs), a complete compendium of temporal information (TM from the acceleration magnitude and components), and the pace at which the activity was performed (FP from  $|\mathbf{a}(t)|$ ). The second best result (90.97%) corresponds to a similar composition ( $F_{22}$ ), which only lacks the latter feature. Finally, it is remarkable the good result (90.02%) obtained with  $F_{20}$  composition, which comprises features extracted from a single signal ( $|\mathbf{a}(t)|$ ). The computational cost of processing a single signal is a clear advantage.



## 4.2. Discussion about other approaches based on cepstral analysis

In the present study, the cepstral coefficients were exclusively extracted from acceleration signals and an exhaustive analysis on the number of required coefficients was performed. In addition, the inclusion of time-domain and moving pace information was also evaluated. As already mentioned, two earlier studies used cepstral analysis for feature extraction. However, those approaches presented some important differences with our study such as they used extra input signals in addition to the acceleration, their data was captured in a laboratory environment, and actual feature vectors are different from ours.

In [30], body and gravitational acceleration, and gyroscope signals were the inputs to the recognizer and they were fused at feature level. This study considered only 3 dynamic activities. Also, it assumed that body and gravitational accelerations can be separated by low-pass filtering from recorded acceleration components. However, components were recorded in directions relative to the orientation of the device, thus being altered by gravity when the device rotates. In addition, proposed features compute frequency band energies using a stretched mel scale. This scale was designed for speech analysis based on the perceptual scale in the human ear, which makes no sense for HAR. In contrast, in the present study cepstral coefficients were computed without any intermediate filter bank. Moreover, body dynamics was explained in terms of cepstral features and a dataset with a vast diversity of dynamic activities was considered.

In a related study [32], the same features proposed in [30] were evaluated for a completely different task: user identification. The authors have found that adapted perceptual linear prediction (PLP) coefficients yielded the lowest error rate for user recognition. Computation of PLP coefficients consider frequency bands defined by a filter bank with a Bark distribution [33], similar to the mel scale discussed before. Since mel and Bark scales were designed for audio signals, it would be interesting to explore the design of a filter bank for acceleration signals. Future works could evaluate evolutionary approaches [41] to obtain an optimized filter bank for activity recognition.

In [31], acceleration and electrocardiogram signals were separately analyzed and fused at decision level. Feature vectors were obtained applying heteroscedastic linear discriminant analysis on normalized cepstral features and their first derivative. Thus, actual features were derived from cepstral analysis but they are not cepstral coefficients. Proposed systems were trained in a subject-dependent fashion, considering 5 subjects in the experiments. By contrast, the present study evaluated subject-independent recognizers, considering 44 subjects in the experiments. Finally, the feature extraction proposed in [31] was applied on too narrow windows (such as 0.48 s) that cannot reveal the quasi-periodic characteristics of human



activities in cepstral domain. Our system used a longer window (approx. 5 s) that can exhibit several periods of an activity, which makes sense to capture that periodicity.

### 4.3. Comparison with previous results on the same dataset

The comparative analysis was performed using the feature vector  $F_{24}$  with 70 CCs, a composition that was selected as the best proposal. This vector comprises 91 features, 76 of which are computed from the acceleration magnitude and the remaining 15 from the acceleration components. The features from acceleration components boost the magnitude features by considering extra information related to axis signals. The comparison will be made with the best system proposed by Xue *et al.* [22], which obtains the best recognition rate reported on the SCUT-NAA dataset. The system uses an SVM as the classifier and a feature vector based on the spectrum of acceleration components. The feature vector has a total of 189 attributes, comprising 63 FFT coefficients obtained from each signal,  $a_x(t)$ ,  $a_y(t)$ , and  $a_z(t)$ . SVM with a radial basis function was used for a fair comparison with previous studies. As in the present proposal, features are extracted from raw acceleration data using a fixed-size sliding window with overlap between consecutive windows.

In Table 6, the proposed system and the reported system are compared by recognition error rate. A comparison with other features is not shown for the sake of space. Features like discrete cosine transform, time-domain features, and autoregressive coefficients were included in [22], with the same experimental setup. The recognition error per activity was highly reduced by the system based on CCs, between 25% and 100%, except for walking and walking quickly. Notably, three activities (jumping, bicycling, and running) were perfectly recognized. The overall recognition error was reduced by 33%. The statistical significance of this reduction was verified with a binomial test performed for the overall recognition rate [42]. Specifically, the binomial test rejected the null hypothesis of both overall rates being equal with a  $p$ -value of 0.016.

The confusion matrix (Table 7) shows that walking and walking quickly were misclassified because the system usually confused between them (low precision and recall). Nevertheless, this confusion is not serious since both activities are expected to be very similar, only differing in their pace. Indeed, an activity that may be considered as walking quickly for one subject may be just normal walking for others. This may explain why the system does not distinguish between them correctly. The remaining activities were classified without important mistakes, as shown by the almost diagonal confusion matrix.

The comparative analysis indicates that the proposed recognition system based on CCs outperformed other reported systems. The proposed feature vector com-

Table 6: Global performance compared by recognition error.

Activity	Xue <i>et al.</i> [22]	$F_{24}$ with 70 CCs	Relative error reduction (%)
Step walking (s)	13.64	<b>2.27</b>	83.36
Jumping (j)	2.33	<b>0.00</b>	100.00
Bicycling (b)	16.67	<b>0.00</b>	100.00
Walking (w)	<b>22.73</b>	25.00	–
Walking backward (wb)	18.18	<b>9.30</b>	48.84
Walking quickly (wq)	<b>18.18</b>	23.81	–
Running (r)	2.27	<b>0.00</b>	100.00
Relaxing (re)	4.55	<b>2.27</b>	50.1
Downstairs (d)	18.18	<b>13.64</b>	24.97
Upstairs (u)	15.91	<b>9.30</b>	41.55
Overall	13.18	<b>8.79</b>	33.31

Table 7: Confusion matrix of feature vector  $F_{24}$  with 70 CCs.

Activity	s	j	b	w	wb	wq	r	re	d	u	Recall
Step walking (s)	43		1								97.7
Jumping (j)		43									100
Bicycling (b)			30								100
Walking (w)				33	2	5			1	3	75.0
W. backward (wb)				2	39	1				1	90.7
W. quickly (wq)				6	1	32			2	1	76.2
Running (r)							44				100
Relaxing (re)			1					43			97.7
Downstairs (d)					1		3		38	2	86.4
Upstairs (u)					1				3	39	90.7
Precision	100	100	93.7	80.5	88.6	84.2	93.6	100	86.4	84.8	

Table 8: Confusion matrix of feature vector  $F_{24}$  with 70 CCs on the walking activities joined.

Activity	s	j	b	w+wq	wb	r	re	d	u	Recall
Step walking (s)	43		1							97.73
Jumping (j)		43								100.0
Bicycling (b)			30							100.0
W. forward (w+wq)				80	1			2	3	93.02
W. backward (wb)				3	39				1	90.70
Running (r)						44				100.0
Relaxing (re)			1				43			97.73
Downstairs (d)					1	3		38	2	86.36
Upstairs (u)				1	1			3	38	88.37
Precision	100	100	93.7	95.2	92.9	93.6	100	88.4	86.4	

bines a variety of information (cepstral, temporal, and pace), therefore it is more robust than a single type of spectral information. Cepstral features give a compact representation of the impulse response (body dynamics) but use a smaller number of coefficients than spectral representations. For instance, 70 CCs are used, in comparison to the 189 FFT coefficients used in [22]. Also, higher recognition rates were obtained by extracting cepstral features from a single signal ( $|\mathbf{a}(t)|$ ), rather than computing frequency-domain features for each of the acceleration components. Adding temporal and pace information prevents confusions based on cepstral information, because they show other aspects of the signals that are not necessarily present in the cepstral or spectral domain. Finally, temporal features from acceleration components add the final piece of information for a minimum extra cost.

#### 4.4. Recognition with walking activities joined

In these experiments, the recognition task was slightly modified. On the assumption that walking and walking quickly are very similar activities, they were joined in a single walking-forward activity. The outputs of previously discussed systems can be post-processed to deal with the new problem, with no need for retraining. If the system has classified a signal as walking or walking quickly, then this signal will be labeled as walking forward. Hence, only 9 activities can be recognized in the new recognition task.

An overall error rate of 10.35% was obtained using the system based on FFT (21.5% of relative error reduction compared to the 10-activity task). The recognition error of walking forward was 6.82% and the recognition errors of the remaining activities were preserved as shown in Table 6. Regarding the system based on CCs and considering the confusion matrix presented in Table 7 the overall error was reduced to 6.18% (29.7% of relative error reduction compared to the 10-activity task). In this case, the recognition error of walking forward was 11.63%. Therefore, when walking and walking quickly are joined, the overall error rate of both systems is significantly reduced.

A different approach for solving the 9-activity task involves building a new classification model, i.e., training a new SVM. Under this scheme, the SVM is trained to recognize 9 activities, including walking forward. This approach assumes that a single model can be a good representation of the walking-forward activity defined above. The results are summarized in the confusion matrix of this approach presented in Table 8. In the 10-activity task (Table 7), signals that correspond exclusively to walking or walking quickly were misclassified 10 times. By contrast, in the 9-activity task, the same signals that correspond to walking forward were misclassified only 6 times (Table 8). This comparison demonstrates that joining the activities is a good strategy. As a global result, the newly trained recognizer for 9 activities and the feature vector  $F_{24}$  with 70 CCs identified walking forward

with a 6.98% error, and it increased precision and recall above 93%. Moreover, the new recognizer reduced the overall error from 6.18% to 5.46%, i.e., a relative error reduction of 11.7% compared to the best 10-activity task recognizer.

## 5. Conclusions

In this study, a human activity recognition system was developed based on the homomorphic analysis of acceleration signals. Body dynamics was captured by a single accelerometer and it was then translated into a compact representation known as cepstrum. The advantage of this representation is that it compresses spectral information while retaining its discriminative power. High recognition rates were achieved with this exclusive representation, which were later improved by the fusion of cepstral, temporal, and pace information at feature level. Error recognition rates were reduced from 100.00% for some activities to 24.97% for others. Therefore, the overall recognition error rate was reduced by 33.31%. The highest recognition levels were achieved with cepstral, temporal, and pace features extracted from acceleration magnitude, combined with temporal features from acceleration components. This is an advantage since acceleration magnitude is known for its independence of the orientation of the recording device (relative to the human body), and the temporal features from acceleration components are not expensive in computational terms.

Future research includes extending the system to work with continuously recorded signals, which requires modeling the long-term dynamics of the system and the automatic segmentation of activities.

## Acknowledgments

This work has been funded by Universidad Nacional del Litoral, PACT CAID 2011-525.

## References

- [1] Y. Meng and H.-C. Kim, «A review of accelerometer-based physical activity measurement», in *Proceedings of the International Conference on IT Convergence and Security 2011*, 120, Springer Netherlands, 2012, pp. 223–237.
- [2] C.-C. Yang and Y.-L. Hsu, «A review of accelerometry-based wearable motion detectors for physical activity monitoring», *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.

- [3] D. Karantonis, M. Narayanan, M. Mathie, N. Lovell, and B. Celler, «Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring», *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 156–167, 2006.
- [4] A. Dalton and G. ÓLaighin, «Comparing supervised learning techniques on the task of physical activity recognition», *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 46–52, 2013.
- [5] R. Poppe, «A survey on vision-based human action recognition», *Image and vision computing*, vol. 28, no. 6, pp. 976–990, 2010.
- [6] A. Nazábal, P. García-Moreno, A. Artés-Rodríguez, and Z. Ghahramani, «Human activity recognition by combining a small number of classifiers», *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1342–1351, 2016.
- [7] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, «Activity classification using realistic data from wearable sensors», *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 119–128, 2006.
- [8] O. Lara and M. Labrador, «A survey on human activity recognition using wearable sensors», *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
- [9] J. Aggarwal and M. Ryoo, «Human activity analysis: A review», *ACM Comput. Surv.*, vol. 43, no. 3, pp. 16:1–16:43, 2011.
- [10] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, «Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions», *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 20–26, 2008.
- [11] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, and Y. Y. Tang, «Person re-identification by dual-regularized kiss metric learning», *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2726–2738, 2016.
- [12] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, «Activity recognition of assembly tasks using body-worn microphones and accelerometers», *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 10, pp. 1553–1567, 2006.
- [13] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon, «Physical activity monitoring based on accelerometry: Validation and comparison with video observation», *Medical and Biological Engineering and Computing*, vol. 37, no. 3, pp. 304–308, 1999.

- [14] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, «Activity identification using body-mounted sensors—a review of classification techniques», *Physiological Measurement*, vol. 30, no. 4, R1, 2009.
- [15] A. Bulling, U. Blanke, and B. Schiele, «A tutorial on human activity recognition using body-worn inertial sensors», *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.
- [16] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, «A survey of online activity recognition using mobile phones», *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015.
- [17] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, «Direct measurement of human movement by accelerometry», *Medical Engineering & Physics*, vol. 30, no. 10, pp. 1364–1386, 2008.
- [18] M. J. Mathie, B. G. Celler, D. N. H. Lovell, and A. C. F. Coster, «Classification of basic daily movements using a triaxial accelerometer», *Medical and Biological Engineering and Computing*, vol. 42, no. 5, pp. 679–687, 2004.
- [19] L. Bao and S. Intille, «Activity recognition from user-annotated acceleration data», *Pervasive computing*, pp. 1–17, 2004.
- [20] R. San-Segundo, J. Lorenzo-Trueba, B. Martínez-González, and J. M. Pardo, «Segmenting human activities based on HMMs using smartphone inertial sensors», *Pervasive and Mobile Computing*, 2016.
- [21] D. Tao, Y. Wen, and R. Hong, «Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition», *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1124–1134, 2016.
- [22] Y. Xue and L. Jin, «A naturalistic 3D acceleration-based activity dataset and benchmark evaluations», in *2010 IEEE International Conference on Systems Man and Cybernetics (SMC)*, Oct. 2010, pp. 4081–4085.
- [23] S. Preece, J. Goulermas, L. Kenney, and D. Howard, «A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data», *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 871–879, 2009.
- [24] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes Ortiz, «Training computationally efficient Smartphone-Based human activity recognition models», in *Artificial Neural Networks and Machine Learning – ICANN 2013*, Springer Berlin Heidelberg, 2013, pp. 426–433.

- [25] Z. He and L. Jin, «Activity recognition from acceleration data based on discrete cosine transform and svm», in *IEEE International Conference on Systems, Man and Cybernetics, 2009*, IEEE, 2009, pp. 5041–5044.
- [26] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. Lee, and J.-S. Wang, «Online classifier construction algorithm for human activity detection using a tri-axial accelerometer», *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 849–860, 2008.
- [27] D. Tao, L. Jin, Y. Yuan, and Y. Xue, «Ensemble manifold rank preserving for acceleration-based human activity recognition», *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1392–1404, 2016.
- [28] Ó. D. Lara, A. J. Pérez, M. A. Labrador, and J. D. Posada, «Centinela: A human activity recognition system based on acceleration and vital sign data», *Pervasive and Mobile Computing*, vol. 8, no. 5, pp. 717–729, 2012.
- [29] A. Oppenheim and R. Schaffer, «From frequency to quefrequency: A history of the cepstrum», *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.
- [30] R. San-Segundo, J. M. Montero, R. Barra-Chicote, F. Fernández, and J. M. Pardo, «Feature extraction from smartphone inertial signals for human activity segmentation», *Signal Processing*, vol. 120, pp. 359–372, 2016.
- [31] M. Li, V. Rozgić, G. Thatté, S. Lee, B. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan, «Multimodal physical activity recognition by fusing temporal and cepstral information», *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, pp. 369–380, 2010.
- [32] R. San-Segundo, R. Cordoba, J. Ferreiros, and L. F. D’Haro-Enríquez, «Frequency features and GMM-UBM approach for gait-based person identification using smartphone inertial signals», *Pattern Recognition Letters*, vol. 73, pp. 60–67, 2016.
- [33] L. R. Rabiner and R. W. Schaffer, *Theory and Applications of Digital Speech Processing*, 1 edition. Upper Saddle River: Prentice Hall, 2011.
- [34] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer New York, 2013.
- [35] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.
- [36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd, ser. Springer Series in Statistics. New York, NY: Springer New York, 2009.

- [37] C.-C. Chang and C.-J. Lin, «LIBSVM: a library for support vector machines», *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 27:1–27:27, 2011.
- [38] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, «The WEKA data mining software: An update», *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.
- [39] H. Liu and L. Yu, «Toward integrating feature selection algorithms for classification and clustering», *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005.
- [40] Y.-W. Chen and C.-J. Lin, «Combining SVMs with various feature selection strategies», in *Feature extraction*, Springer, 2006, pp. 315–324.
- [41] L. D. Vignolo, H. L. Rufiner, D. H. Milone, and J. C. Goddard, «Evolutionary cepstral coefficients», *Applied Soft Computing*, vol. 11, no. 4, pp. 3419–3428, 2011.
- [42] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, 3rd. John Wiley & Sons, 2013.



## Anexo B

# 3D acceleration for heat detection in dairy cows

El artículo fue presentado y ha sido publicado en las memorias del congreso 43° JAIIO, 2014.

S. R. Vanrell, J. O. Chelotti, J. Galli, H. L. Rufiner y D. H. Milone, “3D acceleration for heat detection in dairy cows”, en *Memorias de las 43° JAIIO*, Buenos Aires, Argentina, 2014, págs. 121-134.

Además, debido al impacto tecnológico de las investigaciones y desarrollos implicados, se protegió su propiedad intelectual a partir de la solicitud adjunta en el Anexo [D](#).



# 3D Acceleration for Heat Detection in Dairy Cows

Sebastián R. Vanrell<sup>1,2</sup>, José O. Chelotti<sup>1</sup>, Julio R. Galli<sup>3</sup>, Hugo L. Rufiner<sup>1,2</sup>,  
Diego H. Milone<sup>1,2</sup>

<sup>1</sup> sinc(*i*) – Centro de Investigación en Señales, Sistemas e Inteligencia  
Computacional

Facultad de Ingeniería y Ciencias Hídricas – Universidad Nacional del Litoral

<sup>2</sup> CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

<sup>3</sup> Facultad de Ciencias Agrarias – Universidad Nacional de Rosario

September, 2014

**Abstract.** Accurate and reliable detection of heat in dairy cows is essential for a controlled reproduction and therefore, for maintaining milk production. Classical approaches like visual identification are no longer viable on large dairy herds. Several automated techniques of detection have been proposed, but expected results are only achieved by expensive or invasive methods, because practical methods are not reliable. We present a method that aims to be both practical and accurate. It is based on simple attributes extracted from 3D acceleration data and well known classifiers: multilayer perceptrons, support vector machines and decision trees. Results show promising detection ratios, above 90% in several configurations of the detection system. Best results are achieved with multilayer perceptrons. This information could be readily incorporated to the automated system in a dairy farm and help to improve its efficiency.

**Keywords:** estrus recognition, dairy cattle, binary classification, multilayer perceptron, support vector machine, decision tree, accelerometer

## 1. Introduction

Dairy cows must calve on intervals of 12-14 months to maximize average daily milk production on their lifetime. A deficient detection of heat (estrus) is one of

the main factors that negatively affects calving intervals and reproductive management. Timely insemination is crucial for reproduction and heat detection is the key to the successful use of artificial insemination [1]. The most common cause of a poor detection of heat is failure to watch cows for long enough periods. Cows should be watched at least three times a day, for a period of 20-30 minutes. Thus, as monitoring technology progress, several systems for automated heat detection has been proposed [2]. However, current methods show satisfactory results only for expensive measurements, while simpler methods present high error rates and none of them is completely automatic.

Cows go into heat with an average cycle of 21 days (normal cycle may vary from 18 to 24 days). The average duration of standing heat is 18 hours, and periods between 12 and 24 hours are considered normal. Hot and cold environmental temperatures tend to reduce the length of estrous periods, and increase the difficulty in detecting heat [1]. Successful artificial insemination relies on correct heat identification because it is only possible in a short-time window after heat [3]. Missing a cycle will delay the next calving unnecessarily, causing economical and sustainability complications.

When a cow is in heat exhibits behavioral patterns which are distinctly different from the rest of the estrous cycle. For example, some signs of heat are: increased nervousness and/or restlessness which result in greater physical activity, cows that mount each other, and group formation of sexually active cows [1]. Except for [3], no works on automatic recognition of heat behaviors has been found in the literature. In similar problems, identification of behaviors or activities from sensors has been tried on humans [4], [5] and animals [6]–[10]. For example, authors in [4] attempted to identify everyday human activities with wearable sensors. In [7] is presented an automatic system that decodes ingestive sounds of cattle and gives information of grazing behavior. In [10] a method for behavior recognition of cows using a 3D accelerometer and support vector machine is proposed. Behaviors analyzed —such as standing, feeding, walking— were distinguishable by direct observation of a non expert. However, visual detection of heat behavior is possible but require an expert. Some activities last for minutes and other just for a few seconds ill conditioning the detection problem. Classification shows weak results because some behaviors were often confused.

In the past two decades several methods have been proposed for automatic heat detection [2]. Some of them have excellent accuracy but they involve measurements that are expensive, impractical or invasive. For example, body or milk temperature were considered, but they are not useful for practical applications, since both are highly influenced by external factors [11]. Milk progesterone achieved good rates of detection but at the expense of high costs and extra labor. Voluntary activity of cows were considered as an important indicator of heat, but additional observations

and the use of a cow calendar are suggested as necessary conditions for sufficiently high detection. In recent years, technological improvements have enabled the development of devices that combine activity sensing with other measurements, such as lying/standing conditions. For example, in [3] a change detection algorithm was proposed by combining information from step count and lying time.

Acceleration signals have been found as a feasible source for automated identification of behavior patterns in animals and humans [8], [9], [12], however, no attempts to automatically detect heat from acceleration data have been reported. In this work is presented a strategy that has shown promising results and that has several advantages: data is collected in a simple and not invasive way, by attaching the device to the collar of the animal; recording device and data processor are not expensive; data processing is fast, results can be obtained seconds after that data is downloaded from the device; expert can concentrate only on those cows that show altered behaviors, a useful attribute in large herds. Acceleration records are filtered and segmented. After that, simple statistical attributes are extracted from each segment. Extracted features from a segment are shown to a classifier, which decides if the record belongs to a normal cow or a cow in heat. Tested classifiers are multilayer perceptrons, support vector machines and decision trees.

The rest of this article is organized as follows: Section 2 introduces the analyzed signals and gives a detailed description of the detection method, including feature extraction and classification. Section 3 describes data collection and datasets generation. Results are presented and discussed in Section 4 and finally the conclusions are given in Section 5.

## 2. Detection method

Analyzed signals consist of three-dimensional acceleration varying through time. A fragment of a typical acceleration record is shown in Fig. 1. The three upper plots correspond to the acceleration on north ( $a_n$ ), east ( $a_e$ ), and up ( $a_u$ ) directions. The bottom plot represents the magnitude of the total acceleration vector  $|\mathbf{a}| = \sqrt{a_n^2 + a_e^2 + a_u^2}$ . Accelerations on vertical axes are given in  $\text{m/s}^2$ . Time on horizontal axes is given in s. Visual differences are not evident when signals recorded on normal cows and cows in heat are compared. Thus, a threshold, or some other rule, can not be defined.

A pattern recognition approach where the proposed system learns from labeled examples is applied. In a simplified scheme, the recognition system can be seen as a pipeline with two main stages: feature extraction (Section 2.1) and classification (Section 2.2). The feature extraction stage involves the conversion of the raw signal into a vector of attributes. In the classification stage, these feature vectors are given as an input to the classifier, which estimates the class that each vector

belongs to. In the following subsections both stages will be explained in detail.

## 2.1. Basic feature extraction

The problem of interest is the detection of discriminative changes in the dynamic patterns contained in the records. A record could be extremely long and the translation into only one feature vector would not be the best choice. Furthermore, records of different duration are the typical situation. A common strategy in the field of signal processing involves the division of the original records into time fixed-length segments and then the translation of each segment into a feature vector. The flow of this processing strategy can be seen in Fig. 2. Each feature vector  $\mathbf{x}_i$  represents a segment labeled as  $i$  and is integrated by attributes extracted from  $a_n$ ,  $a_e$ ,  $a_u$  and  $|\mathbf{a}|$ . Therefore, from each record a collection of feature vectors is generated. Five attributes are extracted from the acceleration signal in each direction, which finally are organized to form a feature vector  $\mathbf{x}_i$  of twenty elements.

The shadowed area in Fig. 2 corresponds to the feature extraction detailed for  $a_n$  in Fig. 3. The same is applicable to  $a_u$ ,  $a_e$  and  $|\mathbf{a}|$ . First, a high pass filter is applied to the  $a_n$  segment, thus, baseline components of the signal are separated from those related to movements. The cutoff frequency will be defined later. Then, from the filtered signal five features are extracted: energy, maximum, minimum, amplitude (max – min) and standard deviation. These characteristics are stored as five float numbers that are arranged in a feature vector. Feature extraction processing is developed using Matlab.

In a following stage, feature selection is achieved using methods implemented in [13]. Correlation-based Feature Subset Selection (CFS) is selected for evaluation in combination with an exhaustive search. CFS evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the target class, while having low intercorrelation, are preferred. Restricted by the available data, only a preliminary feature selection is done.

## 2.2. Classification

A proper design of a classifier involves the separation of data into training and testing sets. Each feature vector in the training set contains one target value (the class label) and several attributes (the extracted features). Based on the training data, the goal of the classifier is to predict the target values of the test feature vectors given only the attributes [14].

After successfully representations are constructed from segmented acceleration data, the problem of automatic heat recognition follows naturally. It may be

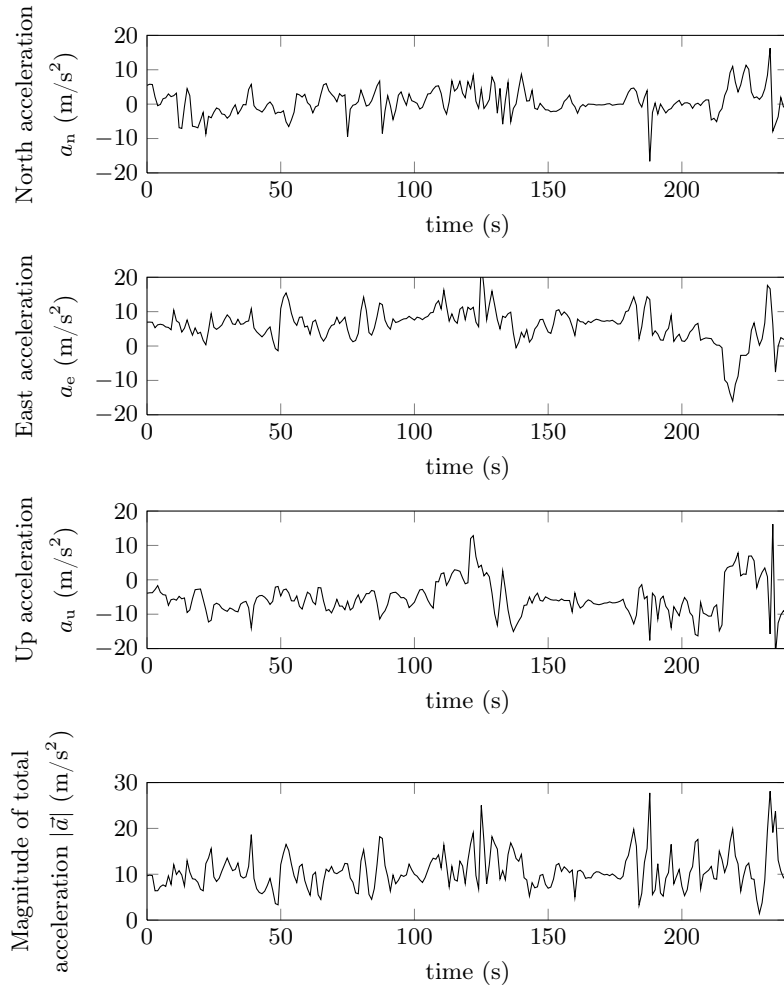


Figure 1: Fragment of a typical record. From top to bottom,  $a_n$ ,  $a_e$  and  $a_u$  plots correspond to the acceleration on north, east and up directions. The magnitude of the total acceleration vector  $|\mathbf{a}| = \sqrt{a_n^2 + a_e^2 + a_u^2}$  is shown in the last plot.

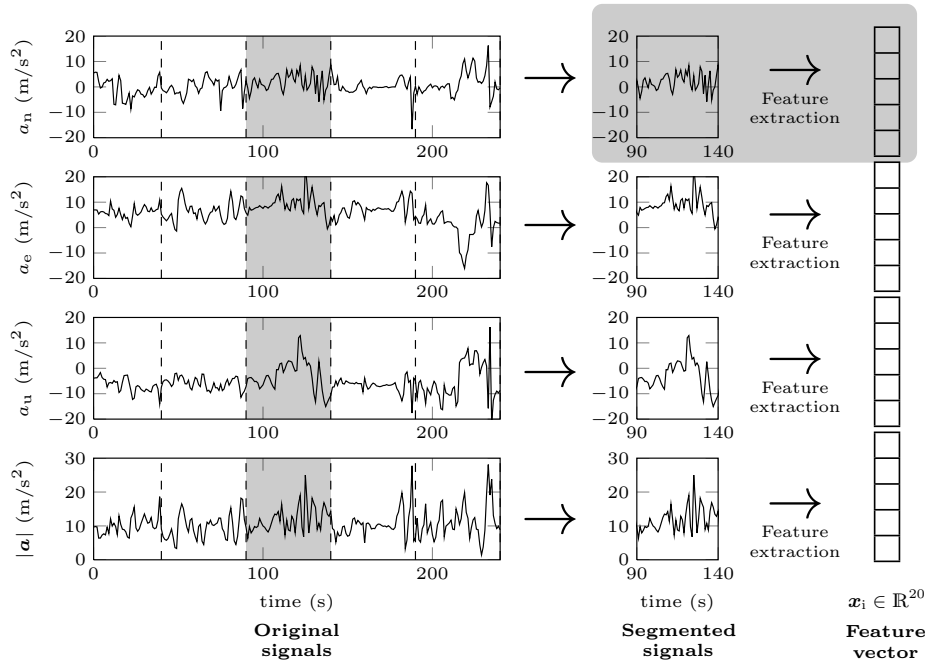


Figure 2: Extraction of feature vector from a segment of the original signal.

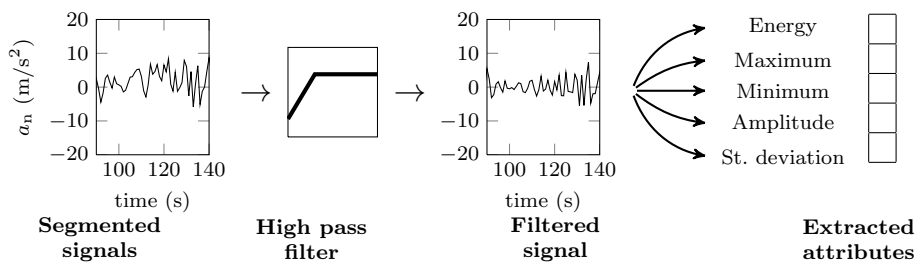


Figure 3: Features extraction from a segment of raw signal.



simply considered as a binary classification problem where each feature vector-label pair consist of 20 attributes and a label that indicates *heat* or *normal*. In this work, three conventional static classifiers are tested: feed forward neural networks, support vector machines, and decision trees.

The assumption of this approach is that each segment contain enough information to represent one of the two target classes. That is, the discriminatory information is conserved after segmentation. Moreover, segments are considered independent. Thus, feature vectors can be presented to classifiers in any order. All classifiers are implemented using [13], except for the support vector machines, that are implemented with [15].

### Multilayer Perceptron

(MLP) is a conventional feed-forward artificial neural network design that can deal with non-linearly separable data. Consists of several layers of nodes in a directed graph, with each layer fully connected to the next one. Each neuron uses a hyperbolic tangent as the activation function. For further details see [16].

In this work, the number of inputs is set to 20 and the number of outputs to 2. One hidden layer is used. Attributes are normalized and labels binarized. Learning rate is fixed to 0.3, momentum to 0.2 and epochs to 1000. The remaining options are leaved in their default values. In preliminary experiments, significant changes were not observed using different number of neurons in the hidden layer. Thus, 5 neurons are used as it provided the best results.

### Support vector machines

(SVM) is a useful machine learning technique for binary classification problems. Conceptually, input feature vectors are non-linearly mapped to a very high-dimension feature space. The mapping to a higher dimension is done by a kernel function  $K$ . In this feature space a linear decision surface is built. That surface is a hyperplane and is located to achieve the maximal margin of separation between the two classes (see [17] or [18] for detailed explanations). In this work, a radial basis function  $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma|\mathbf{x}_i - \mathbf{x}_j|^2}$  is chosen with a soft margin penalty for misclassifications. Penalty coefficient  $C$  and parameter  $\gamma$  are optimized over the independent dataset (sets will be described in Section 3). Remaining parameters of the classifier are kept on their default values.

### Decision Trees

(DT) are characterized by the fact that an unknown feature vector is classified into a class using one or several decision functions in a sequential manner. In

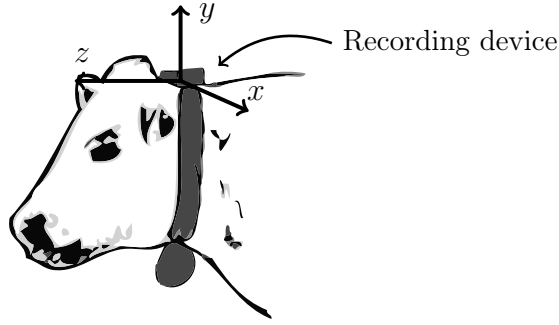


Figure 4: Schematic illustration of recording device attached to the top of the collar. Direction of positive  $x$ ,  $y$  and  $z$  acceleration axes are shown.

general, a decision tree consists of a root node, a number of interior nodes, and a number of terminal nodes. The root node and interior nodes, collectively referred as non-terminal nodes, are linked into decision stages; the terminal nodes represent final classifications. Associated with the root node is the entire set of classes into which a feature vector may be classified. Each node consists of a set of classes to be discriminated, the set of features to be used, and the decision rule for performing the classification. For further details see [19], [20]. In this work, a decision tree is built using the C4.5 algorithm. Changes in pruning confidence did not reflected significant differences. Thus, default settings are used.

### 3. Data collection and datasets

Three-dimensional acceleration were recorded using an inertial measurement unit (IMU) placed inside a box securely attached to the collar of the animal (Fig. 4). The IMU combines triaxial gyroscope, accelerometer, and compass sensors in conjunction with advanced processing to give reliable measurements. Normal signals were recorded during grazing on a cow that was not in heat. Records corresponding to heat were recorded on a cow that exhibited signs of heat. The data collected when the animal was tied or influenced by the experiment itself were excluded from the analysis. Sampling rate was fixed to 1 Hz for all records. The mean record duration was 5 minutes.

Gravity and north directions were registered as versors at the same sampling rate in conjunction with accelerations. East versor was obtained by cross product between gravity and north versors. Up direction was obtained simply as the opposite to gravity. After that, acceleration on  $x$ ,  $y$  and  $z$  axes were projected over north, east and up directions.

In each record, signals were segmented in blocks of 20, 40, 60, 80 and 100

Table 1: Description of datasets generated from raw signals. In the columns: size of segments, number of feature vectors, and cutoff frequency of the filter.

Dataset name	Segments size (seconds)	High pass frequency (Hz)	Number of feature vectors
$D_1$	20	0.00	49
$D_2$	40	0.00	24
$D_3$	60	0.00	15
$D_4$	80	0.00	11
$D_5$	100	0.00	9
$D_6$	20	0.15	49
$D_7$	40	0.15	24
$D_8$	60	0.15	15
$D_9$	80	0.15	11
$D_{10}$	100	0.15	9
$D_{11}$	20	0.25	49
$D_{12}$	40	0.25	24
$D_{13}$	60	0.25	15
$D_{14}$	80	0.25	11
$D_{15}$	100	0.25	9

seconds. The segments, or blocks, were generated by a fixed-length time window, without overlap. The window was moved across the original signals from the end to the beginning, and any remaining samples were discarded. As it was detailed in Section 2.1, feature vectors were created by filtering and extracting features from one segment at a time. The first five datasets listed in Table 1 correspond to the extraction of features without filtering. The remaining ten datasets were obtained by applying a high pass filter with a cutoff frequencies of 0.15 and 0.25 Hz. Those frequencies were selected in preliminary experiments as they seem to give a good balance between energy conservation and low frequencies attenuation.

The resulting number of feature vectors extracted from raw signals are listed in Table 1. Each feature vector belongs to *heat* or *normal* class and is represented by the 20 attributes detailed in Section 2.1

To perform feature and model selection, a subset of 12 feature vectors was extracted from  $D_1$ . Selected feature vectors were removed from  $D_1$  to non invalidate the cross validation of the next section. Therefore, two new sets were defined,  $D_{1s}$  (for selection) and  $D_{1c}$  (for classification), where  $D_{1s} \cup D_{1c} = D_1$  and  $D_{1s} \cap D_{1c} = \emptyset$ . Analogous division was performed over  $D_6$  and  $D_{11}$  sets. Parameter selection for SVM was carried over the union set  $D_{1s} \cup D_{6s} \cup D_{11s}$ .

## 4. Results and discussion

Because the number of feature vectors in each dataset were small, *leave-one-out* cross-validation was applied. In leave-one-out cross-validation the number of folds was fixed to the number of feature vectors in the dataset. In each iteration, one feature vector was used for testing and the remaining feature vectors were used for training the classifier. The test results were collected and averaged over all folds [21]. Performance of classifiers was measured by counting the proportion of correctly predicted feature vectors in the unseen test dataset. That is, the classifier accuracy.

As it was introduced at the end of Section 2.1, a feature selection was carried on. The selected attributes were: total maximum ( $\max_t$ ) in  $D_{1s}$ , up energy ( $\text{ene}_u$ ) in  $D_{6s}$  and east amplitude ( $\text{amp}_e$ ) in  $D_{11s}$ . During classification we compared the performance of using all attributes and only the 3 selected. Datasets that contain all features will be named as  $D_{Af}$ , and datasets that contain only the 3 selected as  $D_{3f}$ .

### 4.1. Detection results

Classification accuracy for the proposed classifiers is shown in Fig. 5. Top plot corresponds to MLP, the middle to SVM and the bottom one to DT. Labels on horizontal axis represent the conditions under each dataset was created. For example, a dataset created by segments of 40 seconds and a filtering frequency of 0.15 Hz is shown as the seventh pair of columns, from left to right. White bars are the result of using all the features and grey bars correspond to accuracy classification using only the selected attributes ( $\max_t$ ,  $\text{ene}_u$  and  $\text{amp}_e$ ).

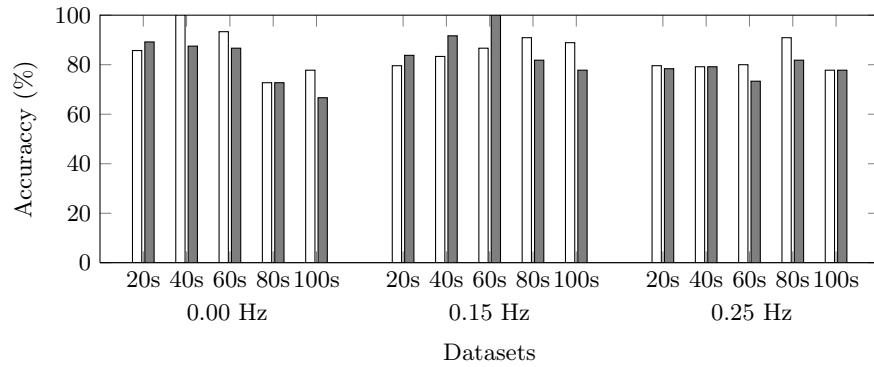
Results for MLP (Fig. 5(a)) using all features are analyzed first. Accuracy achieved over datasets created without filtering ( $D_1$ - $D_5$ ) improves as the length of segments gets reduced. The best result is obtained over the dataset  $D_2$ , that corresponds to segments of 40 seconds length. Classification accuracy on datasets created by filtering at 0.15 Hz ( $D_6$ - $D_{10}$ ) improves as the length of segments is increased. Thus, short segments are preferred for non-filtered signals, and large segments for filtered signals. The reason behind these results is related to the attributes that contribute more to the final decision. Recalling feature selection, total maximum (a local attribute) was selected for non-filtered signals ( $D_{1s}$ ), and up energy (a global measure) was selected for filtered signals at 0.15 Hz ( $D_{6s}$ ). Local attributes are attenuated during filtering process and that explain the apparently opposed results. Finally, accuracy achieved on datasets created by 0.25 Hz filtering frequency ( $D_{11}$  to  $D_{15}$ ) are not influenced by the length of the segments. Because all discriminatory information, that improve results over datasets  $D_1$  to  $D_{10}$ , is no longer available after the 0.25 Hz filtering, in  $D_{11}$  to  $D_{15}$ . Better

results are expected for MLP using selected attributes ( $\max_t$ ,  $\text{ene}_u$  and  $\text{amp}_e$ ), because least and relevant features tend to generalize better than using all features. An improvement can be seen for the datasets that correspond to 0.15 Hz filtering and short segments ( $\leq 60$  s).

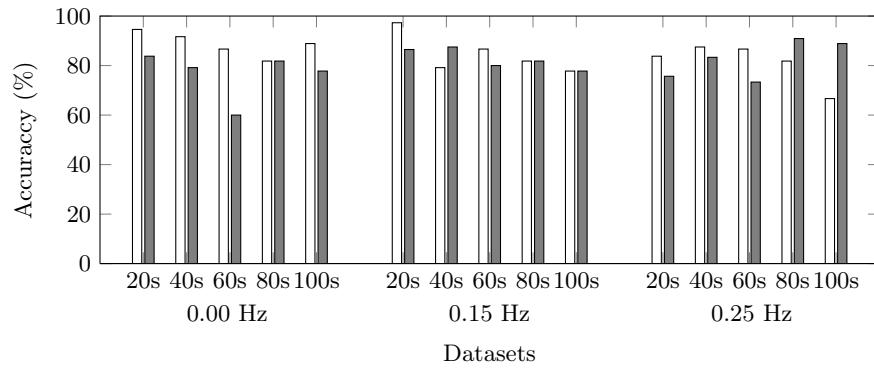
Obtained results using all features and SVM (Fig. 5(b)) were better for short segments, under any filtering condition. This tendency could be influenced by the difficulty to generalize when just a few feature vectors are in the training dataset (the number of feature vectors decreases as length of segments increases). In general, classification using only the selected features show poorer or similar results. The SVM classifier is more immune to irrelevant features, thus, eliminating them do not change the results. Instead, if an important feature is dropped out the performance of the classifier will be reduced. Best result for SVM corresponds to 20 seconds segments, 0.15 Hz filter cutoff and all features.

Classification accuracy for DT (Fig. 5(c)) show results that are reasonably good in almost all datasets, above 77% and reaching a 93% in the best situation. Over datasets created without filtering, a peak is observed on 60 seconds segments and all features. Using only selected features the results improve in the five datasets, and a unique feature is used to build the trees ( $\max_t$ ). The improvements seen on trees built after feature selection are explained by the reduction of the number of possible architectures, a fact that simplify the search. Trees created for datasets corresponding to 0.15 and 0.25 Hz give results that are similar or poorer than previously analyzed.

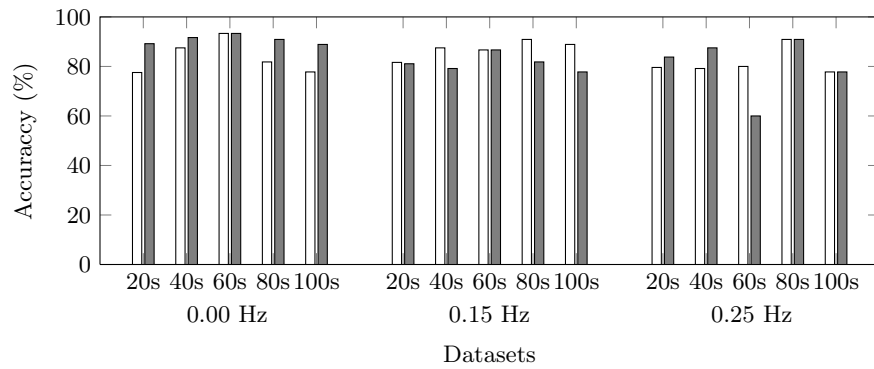
The analysis of the architecture of trees shows that, for datasets in  $D_{Af}$  corresponding to 20 seconds segments, trees look like the one in Fig. 6(a). It is a tree of 7 nodes, 3 non-terminal and 4 terminal (2 for heat and 2 for normal). Threshold values for decisions are given in non-terminal nodes. The remaining trees created (over  $D_{Af}$  and  $D_{3f}$ ) consist of the simplest tree, 1 root node and 2 terminal nodes (1 for each class). An example of such trees is shown in Fig. 6(b). In these simplest trees, the selected feature was total maximum (16 of 27 times). Without feature selection (over  $D_{Af}$ ) the selected feature was north energy (5 of 15 times). DT architecture is interesting as it gives some clues about the classification problem. The classification with only one feature shows that the classes can be distinguish in a simple way. However, it is not always the same feature in the root node. Unexpectedly, north features get that position 10 of 15 times (over  $D_{Af}$ ). Results of classification with trees reveal that minimum and standard deviation are not important attributes to discriminate between classes. An unexpected situation was the absence of features that came from vertical direction, since events with vertical acceleration like mounts are only seen during heat. Total and north direction result as the more relevant ones.



(a) MLP



(b) SVM



(c) DT

Figure 5: Classification accuracy for (a) MLP, (b) SVM and (c) DT. Labels on horizontal axis correspond to the conditions under each dataset was created (size of segments and filtering frequency). White bars were obtained using all the features and gray bars using selected attributes ( $\max_t$ ,  $\text{ene}_u$  and  $\text{amp}_e$ ).

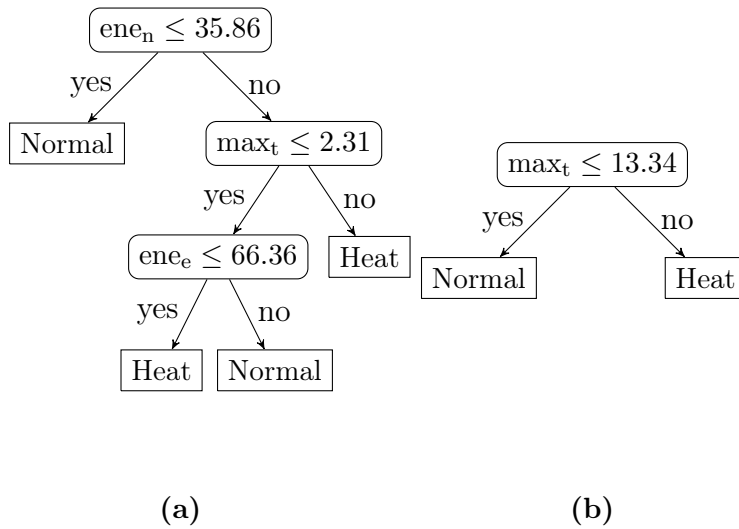


Figure 6: Two examples of trained decision trees. Left tree correspond to  $D_6$  and right tree to  $D_3$ , using all attributes.

## 4.2. Best configuration

The influence of segment size was not always the same when all results are compared. Nevertheless, segments of 40 s and without filtering have lead to good results using any classifier. The results over this dataset are summarized in Table 2, for all classifiers, over  $D_{Af}$  and  $D_{3f}$ . MLP and SVM achieved better results without feature selection, being MLP the best. Instead, DT results were improved after feature selection, followed by MLP. Finally, better results under almost any conditions were achieved with MLP, followed by the performance of DT. SVM showed high variance results, lowering its confidence.

Table 2: Comparative accuracy between classifiers with and without feature selection in  $D_2$  (segments of 40 s, without filtering).

	MLP (%)	SVM (%)	DT (%)
All attributes	100	91.6	87.5
Selected attributes	87.5	79.1	91.6

## 5. Conclusions

A novel approach for heat detection was presented. This is based on commonly used classifiers and simple attributes extracted from acceleration data. The strat-

egy involves two stages, feature extraction —where feature vectors are obtained from filtered and segmented data— and classification —where feature vectors are identified as heat or normal. Results have shown promising levels of detection, turning the proposal into a feasible method for practical applications. To our knowledge this is the first time that accelerations are used in an automatic system of heat recognition. Moreover, previous reported results using non-invasive methods have shown lower accuracy. Our method achieved levels of accuracy that are between 93% and 100%, and it has several advantages: data is gathered in a practical and not invasive way, hardware has low cost, data can be processed in seconds, and an expert can focus only on those cows that are marked as in heat.

Future work will involve the analysis of longer records and the optimization of adjustable parameters of the system.

### **Acknowledgments.**

The authors would like to thank Marcelo Larripa, Alejandra Planisich and Martín Quinteros from Facultad de Ciencias Agrarias, Universidad Nacional de Rosario for their assistance in animal management and gathering data. This work has been funded by Agencia Nacional de Promoción Científica y Tecnológica, Universidad Nacional del Litoral and Universidad Nacional de Rosario, under Projects PICT 2011-2440, PACT CAID 2011 “Señales, Sistemas e Inteligencia Computacional”, CAID 2011-525 and 2013-AGR216.

## **References**

- [1] H. G. Gray and M. A. Varner, «Signs of estrus and improving detection of estrus in cattle», West Virginia University, Fact Sheet IRM 6, 1984.
- [2] R. Firk, E. Stamer, W. Junge, and J. Krieter, «Automation of oestrus detection in dairy cows: A review», *Livestock Production Science*, vol. 75, no. 3, pp. 219–232, 2002.
- [3] R. Jónsson, M. Blanke, N. K. Poulsen, F. Caponetti, and S. Højsgaard, «Oestrus detection in dairy cows from activity and lying data using on-line individual models», *Computers and Electronics in Agriculture*, vol. 76, no. 1, pp. 6–15, 2011.
- [4] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, «Activity classification using realistic data from wearable sensors», *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 119–128, 2006.



- [5] I. Gyllensten and A. Bonomi, «Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life», *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2656–2663, 2011.
- [6] D. Milone, H. Rufiner, J. Galli, E. Laca, and C. Cangiano, «Computational method for segmentation and classification of ingestive sounds in sheep», *Computers and Electronics in Agriculture*, vol. 65, no. 2, pp. 228–237, 2009.
- [7] D. H. Milone, J. R. Galli, C. A. Cangiano, H. L. Rufiner, and E. A. Laca, «Automatic recognition of ingestive sounds of cattle based on hidden Markov models», *Computers and Electronics in Agriculture*, vol. 87, pp. 51–55, 2012.
- [8] W. Hamalainen, M. Jarvinen, P. Martiskainen, and J. Mononen, «Jerk-based feature extraction for robust activity recognition from acceleration data», in *11th International Conference on Intelligent Systems Design and Applications (ISDA)*, Nov. 2011, pp. 831–836.
- [9] R. Nathan, O. Spiegel, S. Fortmann-Roe, R. Harel, M. Wikelski, and W. M. Getz, «Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: General concepts and tools illustrated for griffon vultures», *The Journal of Experimental Biology*, vol. 215, no. 6, pp. 986–996, 2012.
- [10] P. Martiskainen, M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, «Cow behaviour pattern recognition using a three dimensional accelerometer and support vector machines», *Applied Animal Behaviour Science*, vol. 119, no. 1–2, pp. 32–38, 2009.
- [11] D. P. Fordham, T. T. McCarthy, and P. Rowlinson, «An evaluation of milk temperature measurement for detecting oestrus in dairy cattle. II variations in body and milk temperature associated with oestrus», *Veterinary Research Communications*, vol. 11, no. 4, pp. 381–391, 1987.
- [12] A. Reiss, G. Hendebly, and D. Stricker, «A competitive approach for human activity recognition on smartphones», in *ESANN 2013*, ESANN, 2013, pp. 455–460.
- [13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, «The WEKA data mining software: An update», *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.
- [14] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, *A practical guide to support vector classification*, 2003.
- [15] C.-C. Chang and C.-J. Lin, «LIBSVM: a library for support vector machines», *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 27:1–27:27, 2011.

- [16] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006, vol. 1.
- [17] C. Cortes and V. Vapnik, «Support-vector networks», *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [18] C. J. C. Burges, «A tutorial on support vector machines for pattern recognition», *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [19] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. New York: Chapman & Hall, 1993.
- [20] P. H. Swain and H. Hauska, «The decision tree classifier: Design and potential», *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142–147, 1977.
- [21] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical machine learning tools and techniques*. Burlington, Mass.: Morgan Kaufmann, 2011.

## Anexo C

# A regularity-based algorithm for identifying grazing and rumination bouts from acoustic signals in grazing cattle

El artículo ha sido publicado en la revista *Computers and Electronics in Agriculture*.

S. R. Vanrell, J. O. Chelotti, J. R. Galli, S. A. Utsumi, L. L. Giovanini, H. L. Rufiner y D. H. Milone , “A regularity-based algorithm for identifying grazing and rumination bouts from acoustic signals in grazing cattle”, *Computers and Electronics in Agriculture*, vol. 151, págs. 392-402, 2018.

doi: 10.1016/j.compag.2018.06.021



# A regularity-based algorithm for identifying grazing and rumination bouts from acoustic signals in grazing cattle

Sebastián R. Vanrell<sup>1</sup>, José O. Chelotti<sup>1</sup>, Julio R. Galli<sup>2</sup>, Santiago. A. Utsumi<sup>3</sup>, Leonardo L. Giovanini<sup>1</sup>, Hugo L. Rufiner<sup>1,4</sup>, Diego H. Milone<sup>1</sup>

<sup>1</sup> Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL/CONICET, Argentina.

<sup>2</sup> Instituto de Investigaciones en Ciencias Agrarias de Rosario, IICAR, Facultad de Ciencias Agrarias, UNR-CONICET, Argentina.

<sup>3</sup> W.K. Kellogg Biological Station, Michigan State University, United States.

<sup>4</sup> Laboratorio de Cibernética, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina

Received: November 4, 2017; Accepted: June 10, 2018; Published: June 21, 2018

## Abstract.

Continuous monitoring of cattle foraging behavior is a major requirement for precision livestock farming applications. Several strategies have been proposed for this task but monitoring of free-ranging cattle for a long period of time has not been fully achieved yet. In this study, an algorithm is proposed for long-term analysis of foraging behavior that uses the regularity of this behavior to recognize grazing and rumination bouts. Acoustic signals are analyzed offline in two main stages: segmentation and classification. In segmentation, a complete recording is analyzed to detect regular masticatory events and to define the time boundaries of foraging activity blocks. This stage also defines blocks that correspond to no foraging activity (resting bouts). The detection of event regularity is based on the autocorrelation of the sound envelope. For classification, the energy of sound signals within a block is analyzed to detect pauses and to characterize their regularity. Rumination blocks present regular pauses, whereas grazing blocks do not. The evaluation of the proposed algorithm showed very good results for the segmentation task and activity classification. Both tasks were extensively analyzed with a new set of multidimensional metrics. Frame-based F1-score was up to 0.962, 0.891 and 0.935 for segmentation, rumination classification, and grazing classification, respectively. The average time estimation error was below 0.5 min for classification of rumination and grazing on recordings of several hours in length. In addition, a comparison for rumination time estimation was done between the proposed system and a commercial one (Hi-Tag; SCR Engineers Ltd., Netanya, Israel). The proposed algorithm showed a narrower error distribution, with a median of -2.56 min compared to -13.55 min in the commercial system. These results suggest that the proposed system can be used in practical applications.

Web demo available at: <http://sinc.unl.edu.ar/web-demo/rafar/>

**Index terms**— Acoustic monitoring, Activity recognition, Grazing cattle behavior, Precision livestock farming, Signal processing.

# 1. Introduction

In recent years, much effort has been put into the development of animal monitoring applications for precision livestock farming. Monitoring of foraging behavior is key to ensure the fulfillment of the basic health and welfare requirements of grazing cattle and to improve the efficiency of pasture-based production systems [1]. Foraging activities, particularly grazing and rumination, occupy most of the animal's day. Thus, the continuous monitoring of such behavior can help retrieve individual status information for each animal, build a log, detect emerging diseases or the onset of estrus, and optimize pasture and animal management. For example, decreased rumination is interpreted as an indicator of stress [2], anxiety [3], or disease [4]. Conversely, an increase in rumination time is associated with more saliva production and improved rumen health [5].

Cattle foraging behavior is mainly composed of grazing and rumination times. Grazing can cover from 25% to 50% of the day and rumination, from 15% to 40% [6]. The grazing process involves searching, apprehending, chewing, and swallowing herbage. Rumination includes bolus regurgitation, chewing, and deglutition. While grazing, the animal moves its jaw continuously with no predefined interruptions or sequence of events. By contrast, a typical rumination phase involves chewing for 40-60 seconds and a 3-to-5-second interruption during bolus deglutition and regurgitation [1], [7], [8]. During both activities, jaw movements (or masticatory events) are performed rhythmically with a frequency that ranges from 0.75 to 1.20 events per second [9]. The masticatory events are biting, when herbage is apprehended and severed; chewing, when herbage is comminuted; and a compound movement called chew-bite, when herbage is severed and comminuted in the same jaw movement [10]–[12]. Events have a length close to 1 s, whereas activity bouts can last from minutes to hours. Thus, foraging behavior is characterized by events (short timescale) and activities (longer timescale).

Many strategies have been proposed for monitoring foraging behavior, but they are limited by several factors [1], [9], [13]. For instance, foraging behavior could be measured by direct observation or by watching video recordings. However, these methodologies are extremely time-consuming and unfeasible for large herds; besides, it is very difficult to collect data in pasture-based systems over long periods of time. To be of practical use, monitoring should be performed in a fully automatic and noninvasive manner so as not to disturb the normal behavior of the animal. In addition, the system should be capable of working continuously and keep accurate measurements from days to weeks.

Automatic monitoring systems have been developed based on different sensing technologies: motion sensors, noseband pressure sensors, and microphones. The most commonly used motion sensors are accelerometers [14]–[17] and inertial measurement units [18]–[20]. These systems typically seek to recognize a broader set

of activities, such as rumination, grazing, resting, drinking, and walking. An activity is determined by postural analysis of the animal, where the sensors are used to estimate the relative position and motion of its head and body. However, this strategy can confuse activities that share the same posture. For example, resting can be easily confused with rumination, which can be performed while the cow is standing or lying on the ground. A better strategy for recognizing ruminating, eating, and drinking activities is the use of noseband pressure sensors [21]–[25]. The IGER Behavior Recorder was a pioneer development using these sensors. Recently, the RumiWatch system was used to analyze housed and free-ranging cows during one- and two-hour sessions. This yielded very good results, but further studies are required on continuous long-term monitoring. By contrast, acoustic monitoring has proven to be reliable for recognizing short-term ingestive events in free-ranging cows [10], [12], [26]–[31]. A popular monitoring system that includes a logger with a built-in microphone is the Hi-Tag system (SCR Engineers Ltd., Netanya, Israel). However, the sound signal processing is exclusively focused on monitoring rumination in housed cows [32], [33]. No long-term acoustic monitoring of foraging activities has yet been studied for free-ranging cows.

In this study, an algorithm is proposed for identifying grazing, rumination, and resting bouts from acoustic signals. The algorithm provides the start and finish times of each activity block by analyzing the input signal. It is based on the periodic characteristics of jaw movements during grazing and rumination. Jaw-movement sequences, and the occurrence of interruptions, differ greatly between activities. During grazing, bites, chews, and chew-bites are heterogeneously distributed in time with irregular interruptions. Conversely, rumination presents homogeneous phases of chews interrupted by bolus deglutition and regurgitation. The algorithm has two stages. First, the complete recording is analyzed to delimit the blocks of the signal that show periodical jaw movements. The absence of such periodicity defines discarded blocks (resting bouts). Second, the delimited blocks are further analyzed to detect and characterize the interruptions, thus defining which activity corresponds to each block.

The identification of rumination and grazing bouts can be seen as a particular case of continuous activity recognition problem. In this context, recognition systems are typically assessed with standard performance metrics, such as sensitivity, specificity, precision, or correlation coefficient (concordance, Pearson, or Spearman) [24], [25], [34]. However, to use these metrics the problem of continuous activity recognition must be reformulated as a classic classification problem, where input data is mapped to a single category. Unfortunately, restating the problem to conform to standard metrics can be misleading and can produce confusing results [35]. In this study, we propose the use of a new set of multidimensional performance metrics, which provides a detailed description of the recognition pro-

cess at multiple timescales. This allows for a more accurate assessment of the strengths and weaknesses of the proposed recognizers.

## 2. Materials and methods

Grazing and rumination are activities with quasiperiodic characteristics. The proposed *regularity-based algorithm* aims to use this discriminative information to provide grazing and rumination bouts. Two main stages are involved in the offline recognition process: activity segmentation and activity classification (Figure 1). The complete recording is first analyzed to delimit the blocks of the signal that show regular events (jaw movements). A short sliding window on the envelope of the sound signal is used to analyze this regularity. Demarcation of the activity blocks also defines blocks of no activity (resting bouts), which correspond to silence or noisy intervals. Autocorrelation is a well-known technique that has been useful to detect periodicity in noisy signals [36] and it will be used in this stage. During classification, activity blocks are further analyzed to detect interruptions and to characterize their regularity. The energy of the sound signal within a block is analyzed to detect sudden drops, which are related to the interruptions. Regular interruptions are related to bolus deglutition and regurgitation during rumination. Grazing does not show interruptions corresponding to this particular regularity, although it may present irregular interruptions by searching a new plant or patch.

### 2.1. Segmentation by regularity

Segmentation is based on regularity of masticatory events during grazing and rumination. The analysis of the envelope of the sound signal can reveal these events and their periodicity. Envelope computation is the first task of this stage (Figure 1.a). It allows one to operate with low-frequency signals and to discard high-frequency details, unrelated to event regularity [30]. Envelope computation requires three steps: (i) signal rectification, (ii) signal filtering, and (iii) signal subsampling. In the first step the absolute value of signal samples is computed. In the second step the signal is filtered using a low-pass filter, thereby producing the sound envelope. In the third step, a subsample of the original sound envelope is conducted. The main objective of this step is to reduce the computational requirements in the subsequent tasks, since this process significantly reduces the amount of information to be processed without compromising the performance of the algorithm.

Sound envelope is analyzed by frames of 30 s without overlapping (Figure 1.a).



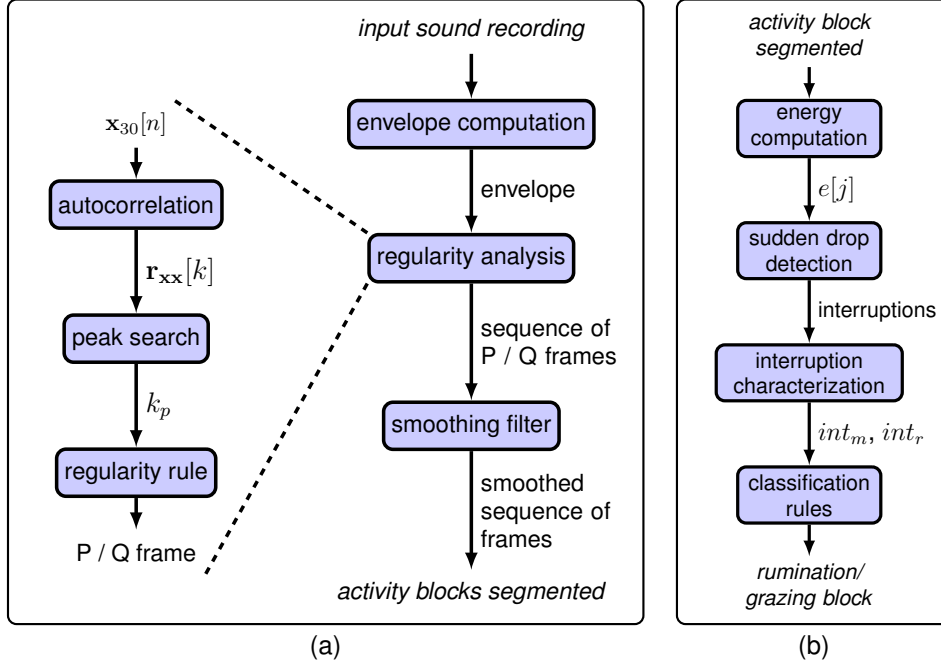


Figure 1: Tasks of a) segmentation and b) classification stages for the proposed algorithm. Steps of regularity analysis during segmentation are also detailed.

The autocorrelation of each frame  $x_{30}[n]$  is performed,

$$r_{xx}[k] = \sum_{n=k}^{N_{30}-1} x_{30}[n]x_{30}[n-k],$$

where  $k \geq 0$  is the lag, and  $N_{30}$  is the number of samples in a frame. Regular activities are expected to have a peak at the typical period of masticatory events. Thus, a local maximum is searched in a surrounding interval  $L_{\text{peak}}$ ,

$$k_p = \arg \max_{k \in L_{\text{peak}}} \{r_{xx}[k]\}.$$

To be considered as a regular activity,  $k_p$  must be in  $L_{\text{reg}}$  (Figure 2). In such a case, a positive label (P) is assigned to the frame. Otherwise, the frame will be set with a negative label (Q), which means that no periodicity was detected.

A sequence of labeled frames (as P or Q) is obtained after the whole recording has been analyzed (Figure 1a). A few intermediate Q frames are undesired breaks from the activity point of view because they lead to fragmentation and unrealistic short activity blocks in the recognized sequence. Since activity segmentation is focused on long-term behavior, these breaks should be reduced. A smoothing filter is applied to labeled sequences to avoid the presence of one or two successive

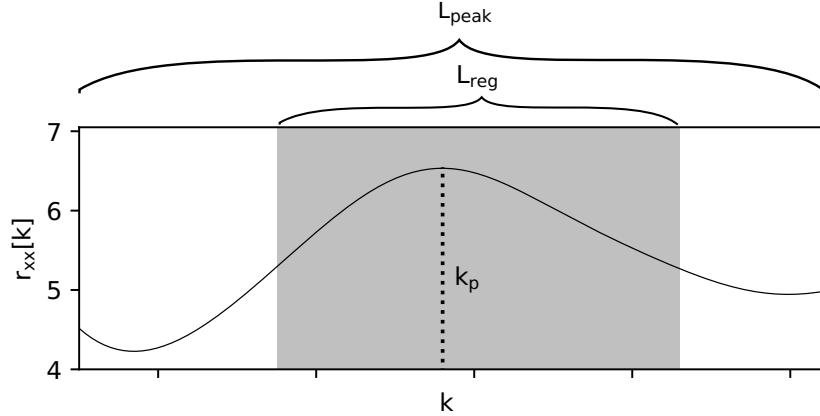


Figure 2: Autocorrelation  $r_{xx}[k]$  of a frame  $x_{30}[n]$  identified as regular. Maximum search in  $L_{\text{peak}}$  yielded  $k_p$  in  $L_{\text{reg}}$ .

Q frames (<1 min) surrounded by P frames. Activity classification is more reliable when long blocks (several minutes) are analyzed.

## 2.2. Classification of activity blocks

Information on event regularity is not enough to discriminate between rumination and grazing. A new characteristic must be extracted to make such distinction. Typical sound waves recorded during grazing and rumination in a free-ranging environment (as detailed in Section 2.4) are shown in Figure 3. It is clear that rumination presents regular interruptions (short periods of low-intensity sound). Grazing does not show interruptions corresponding to this particular regularity, although it may present other interruptions by searching a new plant or patch. Detection of these interruptions and characterizing their regularity are the keys to discriminate between rumination and grazing.

The sound in an activity block is analyzed by 1 s frames  $x_1[n]$  to detect interruptions related to bolus regurgitation (pointed by arrows in Figure 3b). These brief interruptions (3-to-5 s) might be undetected using longer frames, thus 1 s frames were chosen for the analysis. Interruptions could be inferred from changes in the amplitude of the sound signal. However, variations in the amplitude across sessions, microphones, recording devices and cows do not allow a reliable detection. Preliminary experiments showed that an energy-related measure performs better at detecting interruptions. Hence, in this study a frame is characterized by computing the  $e[j] = \log(\|\mathbf{x}_1\|_2)$ , which is proportional to the energy (Figure 1b). The center of a frame  $j$  indicates its position in that block. For example, the

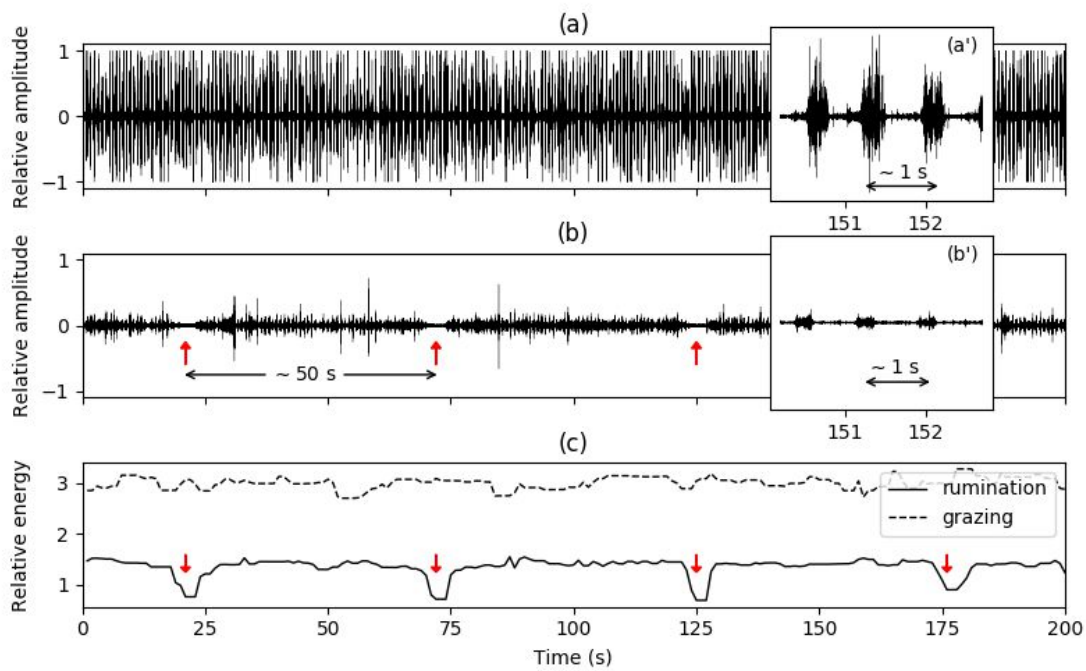


Figure 3: Sound recorded during a) grazing and b) rumination in a free-ranging environment. c) Energy of sound wave during grazing (dashed line), and rumination (solid line). Interruptions related to bolus regurgitation are indicated by arrows. Regularity of events is depicted in a') and b').

solid line in Figure 3c shows the corresponding  $e[j]$  for the sound recorded during rumination (Figure 3b). The algorithm aims to detect sudden drops in  $e[j]$  so as to characterize interruptions.

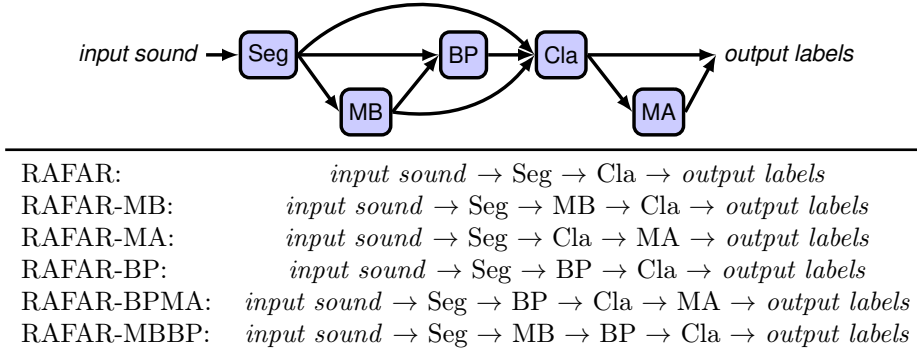
Drop detection starts by computing  $e[j]$  for each 1 s frame in the whole block (Figure 1b). A sliding window of fixed length is used to perform the search. The median  $m_w$  is computed in this window and the first position  $j$ , where  $e[j]$  is lower than a given threshold is assumed to be a sudden drop (to be specified in Section 2.5). The search continues after a big step when a drop is found in the window. Otherwise, a small step is performed. Once drops have been identified, the interruptions in a block are characterized in two ways: (i) the mean interval between interruptions ( $int_m$ ) and (ii) the rate of interruptions per minute ( $int_r$ ). A block would be identified as rumination if the interruptions meet specific criteria. Otherwise, the activity block would be identified as a grazing block.

### 2.3. Processing blocks and gaps

Segmentation by regularity can produce long blocks (hundred of minutes in length) to be classified. These blocks typically correspond to grazing or rumination bouts. However, a single block can occasionally comprise a rumination bout followed by grazing, or vice versa (i.e., a mixed block). Since the classification stage does not change the limits of a block, classifying mixed blocks can introduce partial misclassification. To deal with these misclassifications, larger blocks (>10 min) are analyzed for possible partition. Changes in energy computed by 60 s frames are the guide for partition. A partition is performed if an energy change is greater than a threshold. When this stage is considered, blocks are partitioned before classification. Segmentation can also lead to short gaps (<5 min) between activity blocks, which are not relevant from a practical point of view. These gaps can be merged before or after classification. Prior to classification, short gaps are merged, thus creating long activity blocks to be classified. After merging, the partition of long blocks is expected to play a critical role in system performance. After classification, gap merging is performed only if previous and posterior blocks correspond to the same activity. In this case, short gaps are combined with contiguous blocks to form a long block that is equally labeled. For instance, a short gap between two rumination blocks will be reassigned and a single longer rumination block will be created.

In this study, several variants are evaluated in the combination of these stages of the algorithm. They differ in the order of stages in the process flow. Their names and stages are detailed in Table I. Available stages are segmentation by regularity (Seg), gap merging before classification (MB), partition of long blocks (BP), classification of activity blocks (Cla), and gap merging after classification (MA). The baseline variant is called regularity-based acoustic foraging activity

Table 1: Algorithm variants considered in the experiments. RAFAR: regularity-based acoustic foraging activity recognizer. Seg: segmentation. MB: gap merging before classification. BP: partition of long blocks. Cla: classification of activity blocks. MA: gap merging after classification.



recognizer (RAFAR). Merging stages are mutually exclusive, thus each variant will include MB or MA. The reason is that a gap that would be merged by MA will have already been merged by MB. A detailed description of the parameters of these stages is given in Section 2.5.

## 2.4. Acoustic signal database

The acoustic signals were obtained from an experiment performed at the dairy facility in the Kellogg Biological Station (Michigan State University), in August 2014. Protocols for animal handling and care were reviewed, approved, and conducted according to the Institutional Animal Care and Use Committee of Michigan State University. In this experiment, the foraging behavior of five multiparous Holstein lactating cows grazing perennial ryegrass/white clover and orchardgrass/white clover pastures was continuously monitored during six days. These signals were recorded using SONY ICDPX312 recorders (Figure 4a). A microphone was placed facing inwards on the forehead of cows (Figure 4b) and was protected with rubber foam [29]. All recordings were made at 44.1 kHz sampling rate and 16-bit resolution, providing a nominal 22 kHz recording bandwidth and 96 dB dynamic range, and they were saved in WAV (waveform audio) file format. For this study, 24 hours of recordings containing rumination and grazing sessions were selected to tune the parameters of the algorithm and they were never used again. The results were obtained from 137 hours of recordings, which were selected taking care that they corresponded to a free-ranging environment. Those portions of the recordings that were captured inside the feeding barn were excluded from this study. This selection has been guided by the labels (time-stamps) provided

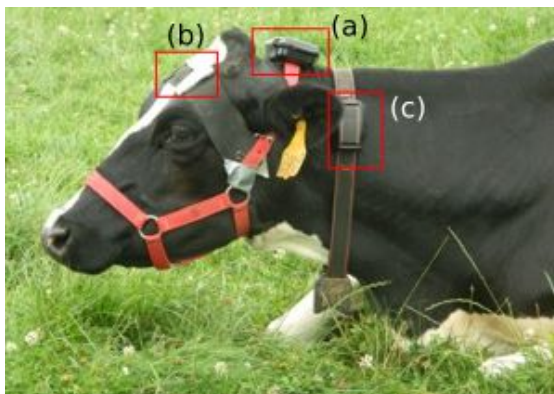


Figure 4: (a) Recording device, (b) microphone and (c) Hi-Tag logger location on the cow’s head and neck.

by the experts.

All the signals used in this study were aurally segmented and labeled independently by two experts in animal behavior, who were able to identify, classify, and label the activity blocks as grazing or rumination. Blocks of no interest were labeled as null. In most cases, experts largely agreed on the labeling of signals, and when there was disagreement, they worked together to reach a final decision. This labeling was used as the reference for comparing and evaluate the performance of the algorithm.

For comparison purposes on rumination time estimation, the Hi-Tag rumination monitor system was used to continuously monitor the animals during the experiments. The Hi-Tag system, consists of rumination loggers, stationary or mobile readers, and software for processing electronic records [32]. Rumination was recorded with this system using a built-in microphone on the collar of the animal (Figure 4.c) and it was summarized as the total time spent ruminating during two-hour chunks.

## 2.5. Experimental setup

The experiments conducted considered the following implementation of the regularity-based algorithm. During segmentation, in the second step of the envelope computation, the signal was filtered using a third-order low-pass Butterworth filter with a cutoff frequency of 2 Hz. In the third step, a subsample of the original sound envelope to 1 kHz was conducted. Sound envelope was analyzed by frames of 30 s and the autocorrelation of each frame  $x_{30}[n]$  was computed. This length provided enough time resolution to catch the periodicity peak in the autocorrelation and the desired time resolution for segmentation. Regular activities were

expected to have a peak at around 0.8 s. This is related to the typical frequency of masticatory events, which is slightly higher than 1 Hz [9]. The local maximum was searched in a surrounding interval  $L_{\text{peak}}$ , which corresponded to (0.3 s, 1.25 s). The frame was considered positive if the maximum was in  $L_{\text{reg}} \subset L_{\text{peak}}$ . This interval corresponded to the interval (0.55 s, 1.06 s) that covered the typical period of masticatory events. The smoothing filter applied to the labeled sequences was a fifth-order median filter. The intervals were defined from experiments with the 24 hour of recordings detailed in Section 2.4. The remaining parameters were defined from preliminary experiments with signals similar to those used in this study.

In the classification stage, the detection of drops in  $e[j]$  was performed with a sliding window of 80 s. This length made possible to find contiguous interruptions during rumination. The median  $m_w$  of energy was computed in this window, which helped to define an adaptive threshold to detect sudden drops in  $e[j]$ . A sudden drop was assumed where  $e[j] < 0.65m_w$ , which has provided the desired sensibility in preliminary experiments. The steps during the search were set to 44 s and 5 s for the big and small step, respectively. These parameters were defined from preliminary experiments with signals similar to those used in this study. The criterion to define an activity block as rumination was that  $int_m \in (25 \text{ s}, 110 \text{ s})$  and  $int_r \in (0.5 \text{ int/min}, 1.5 \text{ int/min})$ . The first condition corresponded to the expected interval between interruptions during rumination and it was computed from detected interruptions. The second one corresponded to the expected frequency of the interruptions, and it considered the number of interruptions and the length of the block. Otherwise, the activity block would be identified as a grazing block. This criterion were defined from experiments with the 24 hour of recordings detailed in Section 2.4.

In the block partition stage, larger blocks (10 min or longer) were analyzed for possible partition. Frames of 60 s were considered to compute the energy and the partition was performed if the relative energy change was greater than a threshold of 0.4. A partition was not performed when the resulting blocks were shorter than 200 s. In the merging stages, a 5 min or shorter gap was considered a short gap that should be merged. These parameters were fixed from experiments with the 24 hour of recordings detailed in Section 2.4. A web demo of the algorithm was developed with the tool [37] and can be accessed at: <http://sinc.unl.edu.ar/web-demo/rafar/>.

## 2.6. Performance metrics

In a standard classification task, a data input is assigned into one, and only one, of the predefined classes. There are clear definitions of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Hence, a recognizer performance can be assessed with standard metrics such as accuracy, precision, re-



call and F1-score [34]. In continuous activity recognition, performance evaluation requires the comparison between a reference sequence and a recognized sequence. The activity blocks of the reference sequence and the recognized sequence may not be in a one-to-one correspondence. For example, a single block of the reference sequence can be partially detected by three shorter blocks in the recognized sequence. The definition of TP does not contemplate this kind of relation between a single input and several outputs. Thus, standard definitions of TP, FP, TN, and FN are no longer suitable. Furthermore, redefining the classification task to provide the input counts required by standard metrics can be equivocal [35]. For instance, standard metrics can be computed based on frames, but they do not distinguish between serious frame errors (block insertion or deletion) and timing offsets (partially detected blocks). Also, they fail to capture common artifacts such as fragmentation, merging, and timing offsets of blocks. A simplified example of the complexity involved in the analysis of continuous activity recognition can be seen in Appendix A

A comprehensive set of performance metrics for continuous activity recognition has been proposed by Ward *et al.* [35]. These metrics are based on two complementary short- and long-term timescales. They present a multidimensional and detailed description instead of a single performance number. In this way, the strengths and weaknesses of a recognizer can be assessed, avoiding ambiguity in the results. Short-term metrics are frame-based, which is the smallest fixed-length unit of time considered by the recognizer. Frame-based metrics facilitate a fine-grain analysis that resembles a continuous time analysis. In this context, a block is defined as a contiguous sequence of equally labeled frames and it has no fixed-length. Long-term metrics are block-based, which provides a different point of view, a big picture of the recognition. This is particularly valuable to detect coarse-grain artifacts and to propose modifications in the recognizer. Errors can be related to segmentation, classification, or a combination of both tasks.

The comparison of two binary sequences is based on the notion of segments. A segment has been defined as the longest part of a block in which the reference and the recognized sequences can be compared in an unambiguous way (i.e., TP, TN, FP, and FN are clearly defined). Segments have no fixed length and can be derived by comparing the reference and the recognized sequences: any change in either sequence marks a segment boundary. This aspect points up a clear difference from blocks and frames, which can be defined from a single sequence. The TP and TN segments keep those labels. The FP and FN segments are assigned into subcategories to better capture block artifacts (details are given in Appendix B). Final labeling of frames is obtained from the corresponding segments. A block is labeled from segments that overlap with them. If a block in the reference sequence is overlapped by a single block in the recognized sequence, the block will be labeled



as C (correctly classified).

The frame- and block-based error metrics were used to characterize each variant of the algorithm. They are false negative rate ( $FNR_*$ ), false discovery rate ( $FDR_*$ ), recall ( $R_*$ ), precision ( $P_*$ ), fragmentation ( $F_*$ ), merging ( $M_*$ ), deletion ( $D_*$ ), insertion ( $I_*$ ), and the standard F1-score ( $F1_*$ ). All metrics were computed for each recording analyzed and then averaged for results presentation. For details about the computation of these metrics see [Appendix C](#).

## 3. Results

### 3.1. Segmentation of foraging activities

Segmentation of foraging activities focuses on the delimitation of activity blocks regardless of their activity label. Three variants of the algorithm were considered: basic segmentation and classification (RAFAR), segmentation and gap merging before classification (RAFAR-MB), and segmentation and gap merging after classification (RAFAR-MA). The other variants in Table 1 were not considered for segmentation analysis because they were included in the selected variants. The block partition (BP) stage in the other variants did not modify the external limits of an activity block, thus segmentation was not altered.

A spider plot considering frame- and block-based error metrics is shown in Figure 5a. A perfect algorithm would yield 0 for each error metric, which matches the boundary of the polygon. Frame-based metrics (left-hand side of the spider plot) revealed low  $FNR$  and  $FDR$  for all variants, which means that segmentation was extremely accurate. The addition of a merging stage before classification (RAFAR-MB) decreased the  $FNR$  and increased the merging errors ( $\sim 13\%$ ). This is expected since gap merging incorporates more positive frames into the recognized sequence. By contrast, gap merging after classification (RAFAR-MA) was more moderate because it merged the gaps that were surrounded by equally classified blocks. Thus, the MA stage had a modest effect compared to the MB stage. In addition, the increase of the merging errors showed a reduction of fragmentation errors.

Regarding block-based metrics (right-hand side of the spider plot), these variants of the algorithm showed surprisingly large  $FNR$  and  $FDR$ . This illustrates the importance of considering long-term error metrics of the sequences, because they can reveal and characterize unseen artifacts of the recognition process. Most of the errors were due to block merging (up to 37%) and block fragmentation (up to 38%). There were also some insertions of blocks. Compared to RAFAR segmentation, the MB stage increased the block merging rate by 17% and reduced block fragmentation by 22%. The MA stage reduced fragmentation and slightly

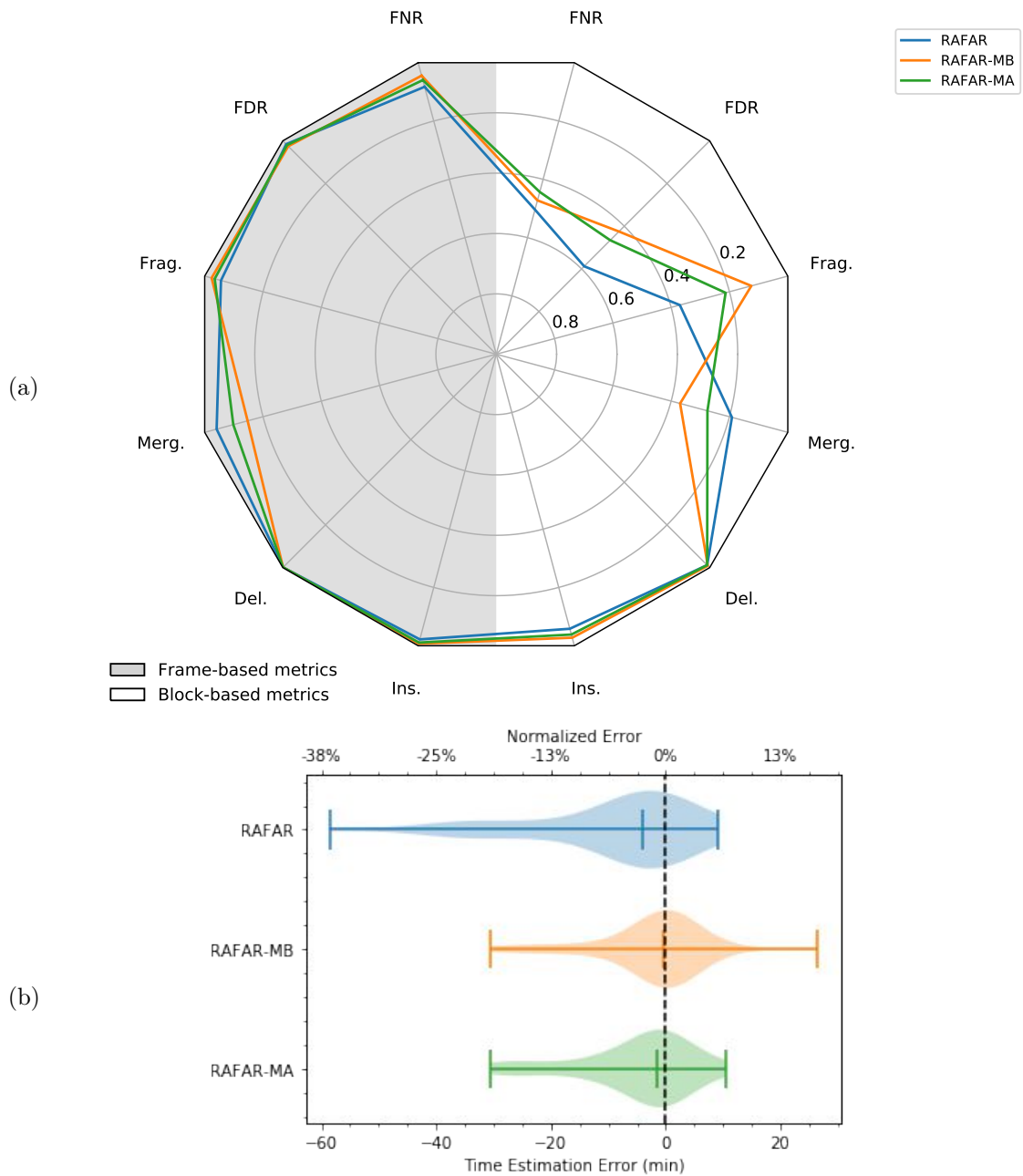


Figure 5: Segmentation of foraging activities. (a) Spider plot of frame- and block-based error metrics. The lowest errors correspond to the boundary of the polygon. (b) Violin plots of time estimation errors. The top axis was normalized with the mean duration of foraging activities (158.6 min) across the recordings analyzed. RAFAR: baseline segmentation and classification. RAFAR-MB: gap merging before classification. RAFAR-MA: gap merging after classification.

increased merging but it had no substantial effect. Block  $FDR$  was reduced but it was still high ( $>40\%$ ). A comparison between frame- and block-based results indicated that most blocks in the recognized sequence correctly overlapped with the blocks of the reference sequence, and most regions of non-foraging activities were not assigned to activity blocks. However, the recognized sequence was mostly comprised of long merging blocks. This should be taken into account for the analysis of classification results and the role of a block partition stage.

A different and practical way of analyzing the segmentation stage was to compare the estimated and the actual duration of the activities. The violin plots in Figure 5b show the time estimation errors in minutes for the variants analyzed. Vertical ticks indicate the median and the extremes of the error distribution. In general, the three variants analyzed achieved very low error with just a few outliers near the extremes. In a single recording, the mean duration of foraging activities was 158.6 min. The algorithm was able to segment these activities with an error ranging between -10 and 10 min. The segmentation with RAFAR tended to underestimate the duration by a median error of 4.16 min. The inclusion of the MB stage in RAFAR-MB substantially reduced the median error of the duration (-0.36 min) but it also showed more overestimation, which is reasonable since the blocks in the recognized sequence are enlarged. Compared to RAFAR, gap merging after classification (RAFAR-MA) reduced the median error to -1.41 min and resulted in a narrower error distribution. This may be explained by an appropriate discrimination of the gaps to merge and an exclusion of the real gaps between activities. It was expected that these accurate segmentation results could help to provide a reliable classification output.

### 3.2. Classification of foraging activities

The results for the classification of foraging activities consider all the RAFAR variants. Baseline classification corresponds to RAFAR, which has the elementary stages in the process flow. Frame-based error metrics showed low  $FNR$  for most RAFAR variants on grazing recognition (Figure 6a). By contrast,  $FDR$  ranged from  $\sim 20\%$  to  $\sim 10\%$ . This means that almost every frame that corresponds to grazing was correctly classified but some RAFAR variants incorrectly identified extra grazing frames (FP frames). The worst variant was RAFAR-MB, which had the highest  $FNR$  and the highest deletion rate. The addition of a block partition stage (RAFAR-BP and RAFAR-BPMA) reduced the  $FDR$  from  $\sim 20\%$  to 15%, which must be related to a decrease in FP frames. Finally, the combination of MB and BP stages (RAFAR-MBBP) achieved the lowest  $FDR$ , that is, a precision above 90% ( $P_f = 1 - FDR_f$ ). Compared to other variants, insertions were highly reduced by RAFAR-MBBP.

Regarding block-based metrics, most RAFAR variants showed a  $FNR$  below

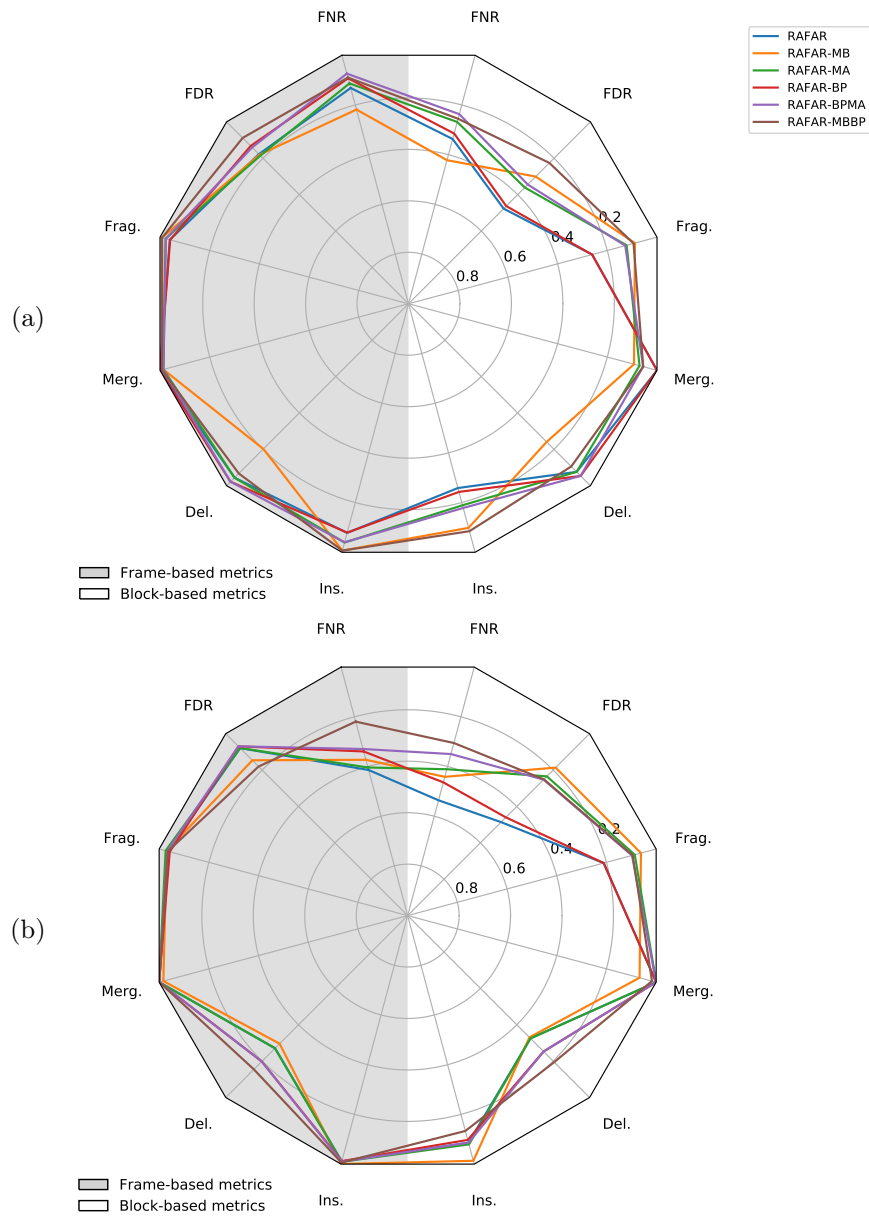


Figure 6: Spider plot of frame- and block-based error metrics for (a) grazing and (b) rumination classification. The lowest errors correspond to the boundary of the polygon. RAFAR: baseline segmentation and classification. RAFAR-MB: gap merging before classification. RAFAR-MA: gap merging after classification. RAFAR-BP: partition of long blocks. RAFAR-BPMA: partition of blocks and gap merging after classification. RAFAR-MBBP: gap merging before classification and partition of blocks.

30% and  $FDR$  below 50% (Figure 6.a). Several grazing blocks of the reference sequence were correctly identified but there were some recognition artifacts. Most errors corresponded to block insertions (up to 26%) and fragmentation (up to 25%). Previous segmentation analyses did not reveal such insertions. Thus, they may be misclassified blocks, which should be compensated by deletion of rumination blocks. RAFAR and RAFAR-BP variants yielded modest results: moderate  $FNR$  and  $FDR$ , 0% merging, 30% fragmentation,  $\sim 25\%$  of insertions and  $<5\%$  deletion. The high fragmentation level indicates that blocks in the reference sequence were partially detected. Therefore,  $FNR$  and  $FDR$  may be increased because of this non critical artifact. RAFAR-MB increased deletions but it significantly reduced fragmentation and insertions as well. A possible explanation is that deletions correspond to blocks identified as rumination after merging grazing and rumination blocks in the MB stage. Fragmentation and insertions were reduced as an expected result of the MB stage. RAFAR-MA and RAFAR-BPMA reduced the insertion and fragmentation errors, which lowered the  $FDR$  by 10% and the  $FNR$  by 5% compared to RAFAR. However, the merging was slightly increased in the MA stage. Finally, the combination of MB and BP stages (RAFAR-MBBP) reached a good compromise, where  $FNR$  and  $FDR$  were close to 20%, and insertions were reduced to less than 10%. Fragmentation, merging and deletion were also limited up to 10%.

The results for rumination recognition are summarized in the spider plot in Figure 6.b. Frame-based metrics showed low  $FDR$  (below 20%) for all RAFAR variants, which means that rumination frames were hardly ever falsely assigned. By contrast, several RAFAR variants identified many rumination frames but not all frames. For instance,  $FNR$  for RAFAR and RAFAR-MA was  $\sim 40\%$  and insertions, merging, and fragmentation were extremely low. The addition of a block partition stage (RAFAR-BP and RAFAR-BPMA) helped to reduce  $FNR$  to  $\sim 30\%$  and deletions to 20%. Once again, the combination of stages in RAFAR-MBBP reached a good compromise between frame  $FNR$  and  $FDR$ , obtaining the lowest deletion rate and very low fragmentation, merging and insertion.

The analysis of block results shows that two variants (RAFAR and RAFAR-BP) correctly recognized  $\sim 50\%$  of the rumination blocks. These variants merged no blocks and presented up to 10% of insertions. The fragmentation errors of 20% indicate that blocks in the reference sequence were partially detected. This explains the moderate  $FNR$  and  $FDR$ . The inclusion of a merging stage (RAFAR-MB or RAFAR-MA) achieved the lowest  $FDR$  and highly reduced fragmentation and insertions. RAFAR-BPMA and RAFAR-MBBP lowered the  $FNR$  and significantly reduced deletions. In addition, these two variants kept fragmentation, merging, and insertions at low rates. The best option was RAFAR-MBBP, which obtained a block  $FNR$  below 30% and a slightly lower  $FDR$ .

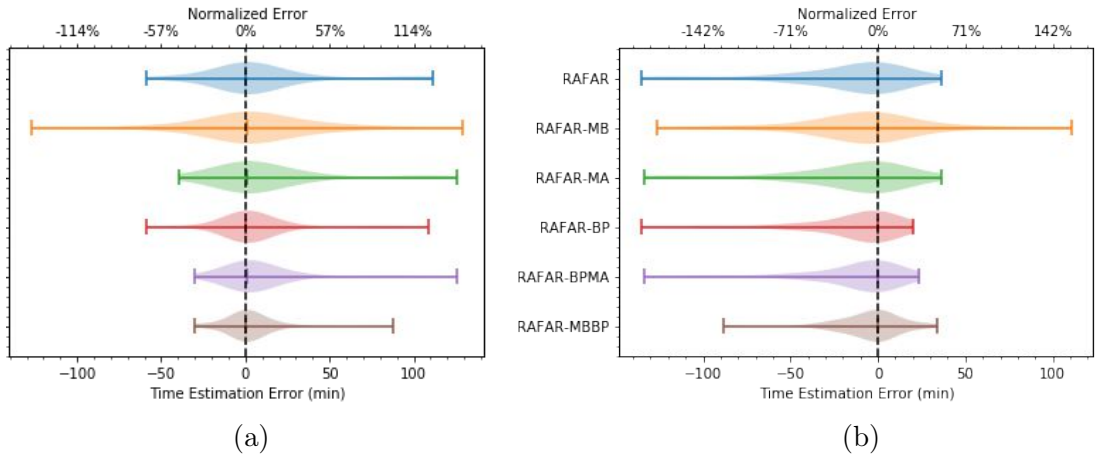


Figure 7: Time estimation error for classification of (a) grazing and (b) rumination. The top axis is normalized with the mean duration across the recordings analyzed for grazing (88.8 min) and rumination (69.9 min). RAFAR: baseline segmentation and classification. RAFAR-MB: gap merging before classification. RAFAR-MA: gap merging after classification. RAFAR-BP: partition of long blocks. RAFAR-BPMA: partition of blocks and gap merging after classification. RAFAR-MBBP: gap merging before classification and partition of blocks.

The results for the estimation of the duration of the foraging activities are shown in the violin plots of Figure 7. They exhibit the error distribution across all the recordings analyzed. In addition, there were a few outliers and the distribution medians of the estimation error were very close to zero for all RAFAR variants. Most RAFAR variants showed a relatively wide distribution, which ranged from -35 min to 30 min. The best results were obtained with RAFAR-MBBP, which had a narrower distribution (-20 to 20 min) and yielded a median estimation error of 0.00 and 0.12 min for rumination and grazing, respectively. These median errors were extremely low compared to the mean duration of the activities in the recordings analyzed: 88.8 min for grazing and 69.9 min for rumination.

## 4. Discussion

The set of metrics proposed for analysis of the algorithm variants provided a new multidimensional view of the recognition performance at two different timescales. Frame-based metrics facilitate a fine-grain analysis that compare the reference and the recognition sequences on a timescale of seconds (1 s frame). Block-based metrics compare the sequences on a timescale of minutes or hours, which provides a big picture of the recognition. To compare and select a variant, a summary

Table 2: Activity segmentation and classification summary. Frame- and block-based F1-score is averaged across signals analyzed (standard deviation).

	Activity segmentation		Rumination classification		Grazing classification	
	Frame-based	Block-based	Frame-based	Block-based	Frame-based	Block-based
RAFAR	0.943 ( $\pm 0.097$ )	0.612 ( $\pm 0.347$ )	0.780 ( $\pm 0.208$ )	0.703 ( $\pm 0.288$ )	0.849 ( $\pm 0.191$ )	0.770 ( $\pm 0.291$ )
RAFAR-MB	<b>0.962</b> ( $\pm 0.057$ )	<b>0.715</b> ( $\pm 0.303$ )	0.778 ( $\pm 0.176$ )	0.818 ( $\pm 0.231$ )	0.878 ( $\pm 0.141$ )	0.829 ( $\pm 0.233$ )
RAFAR-MA	0.956 ( $\pm 0.068$ )	0.688 ( $\pm 0.322$ )	0.789 ( $\pm 0.206$ )	0.791 ( $\pm 0.221$ )	0.854 ( $\pm 0.180$ )	0.787 ( $\pm 0.269$ )
RAFAR-BP	0.943 ( $\pm 0.097$ )	0.612 ( $\pm 0.347$ )	0.836 ( $\pm 0.178$ )	0.719 ( $\pm 0.274$ )	0.882 ( $\pm 0.157$ )	0.770 ( $\pm 0.288$ )
RAFAR-BPMA	0.956 ( $\pm 0.068$ )	0.688 ( $\pm 0.322$ )	0.844 ( $\pm 0.172$ )	0.813 ( $\pm 0.227$ )	0.885 ( $\pm 0.151$ )	0.796 ( $\pm 0.274$ )
RAFAR-MBBP	<b>0.962</b> ( $\pm 0.057$ )	<b>0.715</b> ( $\pm 0.303$ )	<b>0.891</b> ( $\pm 0.125$ )	<b>0.873</b> ( $\pm 0.191$ )	<b>0.935</b> ( $\pm 0.114$ )	<b>0.852</b> ( $\pm 0.225$ )

RAFAR: baseline segmentation and classification. RAFAR-MB: gap merging before classification. RAFAR-MA: gap merging after classification. RAFAR-BP: partition of long blocks. RAFAR-BPMA: partition of long blocks and gap merging after classification. RAFAR-MBBP: gap merging before classification and partition of long blocks.

of segmentation and classification results is given in Table 2. Frame- and block-based results were summarized with the F1-score averaged across the recordings analyzed. Segmentation showed impressive frame-based scores (above 0.94), which means that the proposed algorithm was highly accurate in discriminating foraging activities from others. Block-based scores were up to 0.715. In addition to some blocks being partially detected, most activities were correctly identified. The best activity segmentation was achieved with RAFAR-MB and RAFAR-MBBP (equivalent for segmentation). Regarding classification, frame-based scores were lower than the segmentation for the same algorithm variation, which indicates that some of the correctly identified activity blocks were misclassified in the latter stage. By contrast, block-based scores obtained for classification highly improved segmentation scores. This is particularly evident for RAFAR-MBBP, which combines a merging stage that reduced fragmentation and insertions, and a block partition stage that avoided misclassifications (deletions and insertions exchange between activities). Block fragmentation and insertions were reduced because the merging stage combined the short gaps resulting from the basic segmentation, which resulted in longer blocks for classification. As expected, a single block could include a mix of grazing and rumination bouts. Thus, the partition of blocks helped to prevent the misclassification of mixed blocks. Having analyzed all the variants, RAFAR-MBBP has shown the best tradeoff and will be considered for the following comparisons.



A comparison is made of the rumination time estimation obtained by the Hi-Tag system and the proposed algorithm RAFAR-MBBP. The Hi-Tag rumination system summarizes the total time the animal spent ruminating during two-hour chunks [32]. It provides no access to raw data on timing or duration of rumination bouts within a two-hour chunk [33]. Therefore, the estimations with the RAFAR-MBBP were aligned, and total duration of rumination was summarized to match the same two-hour chunks of the Hi-Tag system. The comparison was made with a total of 53 two-hour chunks from all the recordings analyzed. The remaining 2-hour chunks that overlapped with times of highly-noisy environment (e.g. barn or engine noises) were discarded from this analysis.

The results of time estimation error for rumination are shown in Figure 8. The Hi-Tag system exhibited a wide distribution that ranged from heavy underestimation (-75 min) to equally high overestimation (60 min). In practical terms, these values are not negligible since they are in the same order of magnitude of the two-hour chunk analyzed. The distribution shows a tendency toward underestimation with a median of -13.55 min, in agreement with the results reported in Goldhawk *et al.* [33]. The histogram also shows this tendency and a uniform-like distribution between -40 and 30 min. By contrast, the proposed RAFAR-MBBP obtained a much lower error, which mostly ranged from -25 min of underestimation to 15 min of overestimation. The distribution also shows a tendency toward underestimation with a lower median of -2.56 min. The histogram shows that most errors are very close to zero and that the distribution is clearly narrower. In practical terms, this comparison suggests that the Hi-Tag system estimate rumination time with error that can be greater than 1 hour. By contrast, the proposed system provides rumination time with errors that are only a small time fraction of 1 hour.

The success of the proposed algorithm RAFAR-MBBP in estimating rumination time might be explained by the appropriate combination of two distinctive features of rumination: (i) the regularity of masticatory events and (ii) the homogeneous phases of chews and pauses associated with bolus regurgitation and swallowing. These features allow for the correct discrimination of rumination blocks from grazing blocks. By contrast, the Hi-Tag system might be confusing rumination times with grazing, since it is aimed at detecting rumination exclusively.

It should be noted that the proposed algorithm has been tested with acoustic signals recorded on Holstein dairy cows of similar age and live weight. During the signals recording, cows grazed mixed pastures (ryegrass/white clover and orchard-grass/white clover) in free-ranging environments. Testing the applicability and practicality on other type of animals and grazing environments would require further experiments. For instance, monitoring beef cattle or discontinuous pastures could require an adaptation of the algorithm. The applicability on animals seldom handled by humans could require improvement on the hardware robustness. In dis-



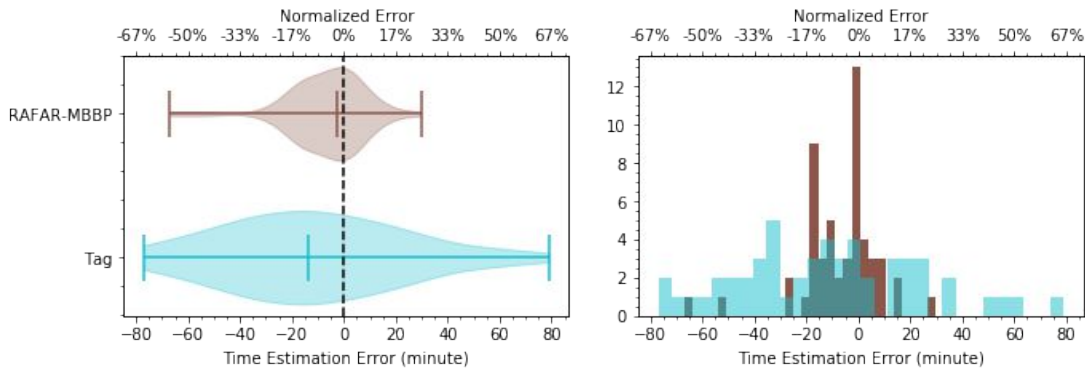


Figure 8: Time estimation error of rumination for RAFAR-MBBP (brown) and Hi-Tag (light blue). The top axis is normalized with the length of segments analyzed (2 hours). RAFAR-MBBP: gap merging before classification and partition of long blocks.

continuous patchy pastures an animal may interrupt its active grazing producing similar interruptions to those sought by the algorithm. However, it is expected that the distinctive regularity of interruptions during rumination will continue to allow discriminating rumination from grazing, even in patchy pastures.

The design of the algorithm has been focused on offline processing of acoustic recordings of several hours. It provides long-term timescale analysis of ruminant foraging behavior, and it complements previous acoustic methods that have been suitable for the recognition and characterization of jaw movements on a short-term timescale. The results were highly satisfactory since sound recordings of several hours were processed in a few minutes on a standard desktop computer (e.g., 6 h of sound recording can be processed in approximately 5 min). It is probable that the proposed equipment is less practical than the Hi-Tag system. The Hi-Tag system has been designed for commercial purposes (simple and easy to apply in a wide range of environments). For example, the attachment procedure is straightforward and it can operate for long periods of time without being recharged. The hardware of our system requires further development to provide ease of use and flexibility in diverse environments, to take advantage of its superior performance in a wider range of applications.

## 5. Conclusions

In this study, an algorithm is proposed for segmenting and classifying foraging activity bouts in grazing cows. Remarkable results were obtained for segmentation. Frame-based F1-score was above 0.96 and the average time estimation error was

below 0.5 min for signals of several hours. Classification of rumination and grazing also produced very good results. For both activities, the frame-based F1-score was above 0.89 and the average time estimation error was very close to zero. Grazing was slightly better identified than rumination. The proposed system estimated rumination duration with a much lower error than the commercial system on a free-ranging environment. The median of the error were -2.56 min for the proposed system and -13.55 min for the commercial one. The proposed system can be implemented for practical applications on similar environments to those considered in this study. Future work will focus on testing the algorithm on other types of animals and grazing environments, on developing a preprocessing stage to deal with highly noisy environments, and on applying advanced machine learning techniques to improve classification.

## Acknowledgements

The authors would like to thank Walter Schulte from Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral for his work as an undergraduate student during the early stages of this project. This study has been funded by Universidad Nacional del Litoral, PACT CAID 2011-525, Universidad Nacional de Rosario, project 2013-AGR216 and Agencia Santafesina de Ciencia, Tecnología e Innovación (ASACTEI), project 2010-216-16 (2016). The method presented within this publication is protected under the international patent application PCT/IB2015/053721.

## Appendix A. Example of continuous activity recognition

A simplified example is presented in Figure [A.1](#) to show the complexity involved in the analysis of continuous activity recognition. Two recognized sequences (A and B) are compared with a reference sequence. Recognizer A returns a sequence that tends to merge blocks of the reference, whereas recognizer B yields a much more fragmented sequence compared to the reference. The number of blocks in each sequence is different. The first block of the reference is fragmented in sequence B. The second and third blocks are merged in sequence A. Then, there are an insertion, a deletion, and the last two blocks are simultaneously fragmented and merged in sequence B. This kind of visual analysis can provide insight into the recognition process. However, it is difficult to establish which recognizer is better.

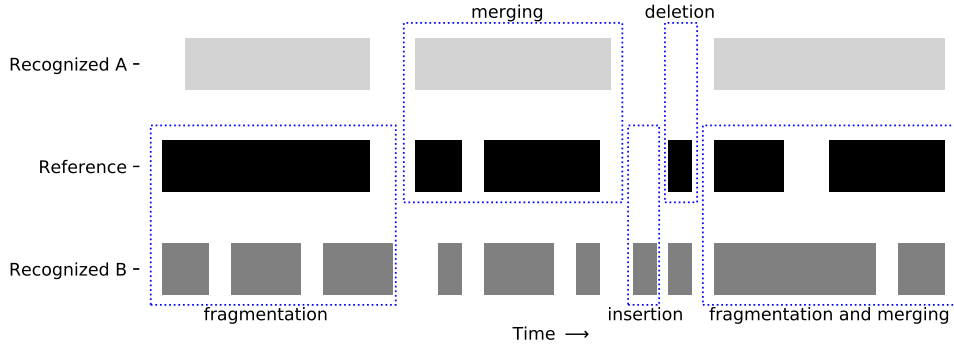


Figure A.1: Example of recognized sequences A and B compared to the reference sequence. Common artifacts found in continuous activity recognition are surrounded by dashed blue rectangles.

## Appendix B. Labeling of segments

The TP and TN segments keep those labels. The FP and FN segments are assigned into subcategories to better capture block artifacts. The subcategories are summarized in Table B.1 and an example of segment assignments is shown in Figure B.1.

Table B.1: Subclassification of false positive (FP) and false negative (FN) segments.

FP	I (insertion)	a FP that corresponds to an isolated insertion in the recognized sequence
	M (merge)	a FP that occurs between two TP segments of a merged block
	O (overfill)	a FP that occurs at the start or the end of a partially matched block
FN	D (deletion)	a FN that corresponds to an isolated deletion in the recognized sequence
	F (fragmentation)	a FN that occurs between two TP segments of a fragmented block
	U (underfill)	a FN that occurs at the start or the end of a partially matched block

## Appendix C. Definitions of frame- and block-based error metrics

The frame- and block-based error metrics are defined in Table C.1. Frame-based metrics are defined considering the counts of true positives  $TP$ , false positives  $FP$ , false negatives  $FN$ , fragmented  $F$ , merged  $M$ , and deleted  $D$  frames in the reference sequence, and the count of inserted  $I$  frames in the recognized sequence, respectively. Frames of 1 s were considered as the smallest time unit

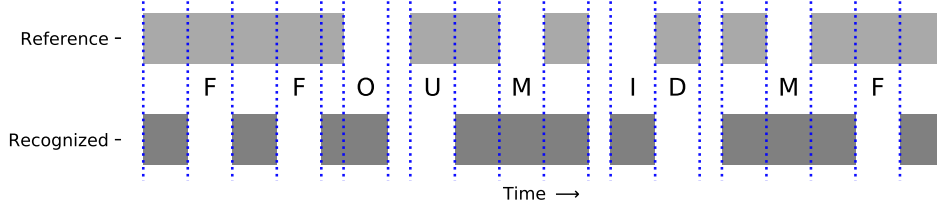


Figure B.1: Segments and boundaries defined from the comparison between the reference and the recognized sequences. Segments that would be usually classified as FN or FP are identified with the corresponding subcategory. The TP and TN labels are omitted for clarity.

Table C.1: Definitions of frame- and block-based error metrics.

Error metric	Frame-based	Block-based
False negative rate	$FN R_f = 1 - \frac{TP}{TP+FN} = 1 - R_f$	$FN R_b = 1 - \frac{C}{B^{ref}} = 1 - R_b$
False discovery rate	$FDR_f = 1 - \frac{TP}{TP+FP} = 1 - P_f$	$FDR_b = 1 - \frac{C}{B^{rec}} = 1 - P_b$
Fragmentation	$F_f = \frac{F}{TP+FP}$	$F_b = \frac{F}{B^{ref}}$
Merging	$M_f = \frac{M}{TP+FP}$	$M_b = \frac{M}{B^{ref}}$
Deletion	$D_f = \frac{D}{TP+FP}$	$D_b = \frac{D}{B^{ref}}$
Insertion	$I_f = \frac{I}{TP+FN}$	$I_b = \frac{I}{B^{rec}}$

for results analysis. Block-based metrics are defined considering the counts of total ( $B^{ref}$ ), correctly detected ( $C$ ), fragmented ( $F$ ), merged ( $M$ ), and deleted ( $D$ ) blocks in the reference sequence, and the counts of total ( $B^{rec}$ ) and inserted ( $I$ ) blocks in the recognized sequence, respectively. In addition, the standard F1-score was computed for frames  $F1_f = \frac{2R_f P_f}{R_f + P_f}$  and blocks  $F1_b = \frac{2R_b P_b}{R_b + P_b}$  based on the corresponding precision and recall defined in Table [C.1](#).

## References

- [1] J. Hodgson and A. W. Illius, «The ecology and management of grazing systems», 1998.
- [2] M. S. Herskin, L. Munksgaard, and J. Ladewig, «Effects of acute stressors on nociception, adrenocortical responses and behavior of dairy cows», en, *Physiol. Behav.*, vol. 83, no. 3, pp. 411–420, Dec. 2004.
- [3] D. J. Bristow and D. S. Holmes, «Cortisol levels and anxiety-related behaviors in cattle», *Physiol. Behav.*, vol. 90, no. 4, pp. 626–628, Mar. 2007.

- [4] J. G. Welch, «Rumination, particle size and passage from the rumen», *J. Anim. Sci.*, vol. 54, no. 4, pp. 885–894, 1982.
- [5] K. A. Beauchemin, «Ingestion and mastication of feed by dairy cattle», *Vet. Clin. North Am. Food Anim. Pract.*, vol. 7, no. 2, pp. 439–463, 1991.
- [6] R. J. Kilgour, «In pursuit of “normal”: A review of the behaviour of cattle at pasture», *Appl. Anim. Behav. Sci.*, vol. 138, no. 1, pp. 1–11, 2012.
- [7] J. K. D. Trindade, P. C. d. F. Carvalho, F. P. Neves, C. E. Pinto, H. L. Gonda, L. B. Nadin, and L. H. S. Correia, «Potencial de um método acústico em quantificar as atividades de bovinos em pastejo», *Pesqui. Agropecu. Bras.*, vol. 46, no. 8, pp. 965–968, 2011.
- [8] M. A. Benvenuti, D. R. Pavetti, D. P. Poppi, I. J. Gordon, and C. A. Cangiano, «Defoliation patterns and their implications for the management of vegetative tropical pastures to control intake and diet quality by cattle», *Grass Forage Sci.*, vol. 71, no. 3, pp. 424–436, Sep. 2016.
- [9] A. Andriamandroso, J. Bindelle, B. Mercatoris, and F. Lebeau, «A review on the use of sensors to monitor cattle jaw movements and behavior when grazing», *Biotechnology, Agronomy, Society and Environment*, vol. 20, 2016.
- [10] E. A. Laca, E. D. Ungar, N. G. Seligman, M. R. Ramey, and M. W. Dement, «An integrated methodology for studying short-term grazing behaviour of cattle», *Grass Forage Sci.*, vol. 47, no. 1, pp. 81–90, 1992.
- [11] E. D. Ungar and S. M. Rutter, «Classifying cattle jaw movements: Comparing IGER behaviour recorder and acoustic techniques», *Appl. Anim. Behav. Sci.*, vol. 98, no. 1, pp. 11–27, 2006.
- [12] J. R. Galli, C. A. Cangiano, M. A. Pece, M. J. Larripa, D. H. Milone, S. A. Utsumi, and E. A. Laca, «Monitoring and assessment of ingestive chewing sounds for prediction of herbage intake rate in grazing cattle», *Animal: An International Journal of Animal Bioscience*, pp. 1–10, 2017.
- [13] R. Delagarde, J.-P. Caudal, and J.-L. Peyraud, «Development of an automatic bitemeter for grazing cattle», in *Annales de zootechnie*, vol. 48, 1999, pp. 329–339.
- [14] L. A. González, G. J. Bishop-Hurley, R. N. Handcock, and C. Crossman, «Behavioral classification of data from collars containing motion sensors in grazing cattle», *Comput. Electron. Agric.*, vol. 110, pp. 91–102, 2015.
- [15] C. Arcidiacono, S. M. C. Porto, M. Mancino, and G. Cascone, «Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data», *Comput. Electron. Agric.*, vol. 134, pp. 124–134, 2017.

- [16] V. Giovanetti, M. Decandia, G. Molle, M. Acciaro, M. Mameli, A. Cabiddu, R. Cossu, M. G. Serra, C. Manca, S. P. G. Rassu, and C. Dimauro, «Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer», *Livest. Sci.*, vol. 196, pp. 42–48, 2017.
- [17] P. Martiskainen, M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, «Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines», *Appl. Anim. Behav. Sci.*, vol. 119, no. 1–2, pp. 32–38, 2009.
- [18] A. L. H. Andriamandroso, F. Lebeau, Y. Beckers, E. Froidmont, I. Dufrasne, B. Heinesch, P. Dumortier, G. Blanchy, Y. Blaise, and J. Bindelle, «Development of an open-source algorithm based on inertial measurement units (IMU) of a smartphone to detect cattle grass intake and ruminating behaviors», *Comput. Electron. Agric.*, vol. 139, pp. 126–137, 2017.
- [19] D. Smith, A. Rahman, G. J. Bishop-Hurley, J. Hills, S. Shahriar, D. Henry, and R. Rawnsley, «Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems», *Comput. Electron. Agric.*, vol. 131, pp. 40–50, 2016.
- [20] P. L. Greenwood, D. R. Paull, J. McNally, T. Kalinowski, D. Ebert, B. Little, D. V. Smith, A. Rahman, P. Valencia, A. B. Ingham, and G. J. Bishop-Hurley, «Use of sensor-determined behaviours to develop algorithms for pasture intake by individual grazing cattle», *Crop Pasture Sci.*, vol. 68, no. 12, pp. 1091–1099, 2018.
- [21] S. Rutter, R. Champion, and P. Penning, «An automatic system to record foraging behaviour in free-ranging ruminants», *Applied Animal Behaviour Science*, vol. 54, no. 2-3, pp. 185–195, 1997.
- [22] S. M. Rutter, «Graze: A program to analyze recordings of the jaw movements of ruminants», *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, vol. 32, no. 1, pp. 86–92, 2000.
- [23] F. Nydegger, L. Gyga, and W. Egli, «Automatic measurement of rumination and feeding activity using a pressure sensor.», *International Conference on Agricultural Engineering - AgEng 2010: Towards environmental technologies, Clermont-Ferrand, France, 6-8 September 2010*, 2010.
- [24] N. Zehner, C. Umstätter, J. J. Niederhauser, and M. Schick, «System specification and validation of a noseband pressure sensor for measurement of ruminating and eating behavior in stable-fed cows», *Comput. Electron. Agric.*, vol. 136, pp. 31–41, 2017.

- [25] J. Werner, L. Leso, C. Umstatter, J. Niederhauser, E. Kennedy, A. Geoghegan, L. Shalloo, M. Schick, and B. O'Brien, «Evaluation of the RumiWatch-System for measuring grazing behaviour of cows», *Journal of Neuroscience Methods*, vol. 300, pp. 138–146, 2018.
- [26] J. R. Galli, C. A. Cangiano, D. H. Milone, and E. A. Laca, «Acoustic monitoring of short-term ingestive behavior and intake in grazing sheep», *Livestock Science*, vol. 140, no. 1–3, pp. 32–41, 2011.
- [27] W. M. Clapham, J. M. Fedders, K. Beeman, and J. P. S. Neel, «Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle», *Comput. Electron. Agric.*, vol. 76, no. 1, pp. 96–104, 2011.
- [28] S. Navon, A. Mizrach, A. Hetzroni, and E. D. Ungar, «Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring», *Biosystems Eng.*, vol. 114, no. 4, pp. 474–483, 2013.
- [29] D. H. Milone, J. R. Galli, C. A. Cangiano, H. L. Rufiner, and E. A. Laca, «Automatic recognition of ingestive sounds of cattle based on hidden Markov models», *Comput. Electron. Agric.*, vol. 87, pp. 51–55, 2012.
- [30] J. O. Chelotti, S. R. Vanrell, D. H. Milone, S. A. Utsumi, J. R. Galli, H. Leonardo Rufiner, and L. L. Giovanini, «A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle», *Comput. Electron. Agric.*, vol. 127, pp. 64–75, 2016.
- [31] J. O. Chelotti, S. R. Vanrell, J. R. Galli, L. L. Giovanini, and H. L. Rufiner, «A pattern recognition approach for detecting and classifying jaw movements in grazing cattle», *Computers and Electronics in Agriculture*, vol. 145, pp. 83–91, Feb. 2018.
- [32] K. Schirmann, M. A. G. von Keyserlingk, D. M. Weary, D. M. Veira, and W. Heuwieser, «Validation of a system for monitoring rumination in dairy cows», *J. Dairy Sci.*, vol. 92, no. 12, pp. 6052–6055, 2009.
- [33] C. Goldhawk, K. Schwartzkopf-Genswein, and K. A. Beauchemin, «Validation of rumination collars for beef cattle», *J. Anim. Sci.*, vol. 91, no. 6, pp. 2858–2862, 2013.
- [34] M. Sokolova and G. Lapalme, «A systematic analysis of performance measures for classification tasks», *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.
- [35] J. A. Ward, P. Lukowicz, and H. W. Gellersen, «Performance metrics for activity recognition», *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 1, 6:1–6:23, Jan. 2011.

- [36] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Pearson Education, 2011.
- [37] G. Stegmayer, M. Pividori, and D. H. Milone, «A very simple and fast way to access and validate algorithms in reproducible research», *Briefings in Bioinformatics*, vol. 17, no. 1, pp. 180–183, 2016.



## Anexo D

# Proceso para detectar celo en un animal rumiante y dispositivo

La siguiente solicitud de patente fue presentada en el Instituto Nacional de la Propiedad Industrial de la República Argentina:

Patente nacional. Inventores: J. R. Galli, L. L. Giovanini, D. H. Milone, H. L. Rufiner, J. O. Chelotti y S. R. Vanrell, “Proceso para detectar celo en un animal rumiante y dispositivo”. Titulares: UNL, CONICET, UNR. Expediente: P20160103881. Fecha de aplicación: 2016-12-16



# **Sistema para el monitoreo y evaluación de actividades reproductivas realizadas por rumiantes**

## ***Campo de la invención***

La presente invención proporciona un método y un dispositivo no invasivo para detectar el celo en bovinos mediante la detección, identificación y análisis, a lo largo de tiempo, de los cambios de comportamiento alimentario, grupal y físico a partir del registro y análisis de los movimientos, posición y sonidos producidos por el animal durante su alimentación y su relación con los períodos de celo.

## ***Estado de la Técnica***

La competitividad de la ganadería (carne y leche) depende en gran medida del desempeño reproductivo del rodeo. Básicamente, las vacas (tanto para producción de terneros o de leche) deberían tener un parto todos los años. Es por eso que se considera relevante mejorar la utilización de los recursos disponibles a lo largo de todo el ciclo para aumentar la eficiencia reproductiva de los rodeos. Cada vaca que no queda preñada, se convierte en un recurso no productivo que debe mantenerse y en consecuencia cae la eficiencia productiva. Por lo tanto, un manejo reproductivo eficaz es determinante en la productividad de los sistemas ganaderos. Para ello, es importante determinar cuándo una vaca está en celo, ya que esto permite determinar el momento adecuado para la monta natural por los toros (servicio a corral) o la inseminación artificial por los operarios.

Para una eficiente detección de celos, lo más importante es conocer los síntomas que se manifiestan. El síntoma más confiable de una vaca en celo es que se deja montar (permance inmóvil) por los toros u otras vacas. Este síntoma característico puede durar de cuatro a seis segundos. Aproximadamente, la frecuencia de las montas de una vaca en celo es de tres o cuatro veces cada dos horas, y este comportamiento (celo o estro) es visible durante seis a doce horas, dependiendo de varios factores del animal y del ambiente (época el año).

Además de dejarse montar (inmovilidad), algunos otros síntomas de la vaca en celo son:

- Monta a otras vacas del rodeo;
- Se observan descargas vaginales de mucosas claras;
- Apoya la quijada sobre la grupa de otras vacas;
- Se ve la vulva hinchada y enrojecida;
- Se presentan los flancos sucios y cola raspada en su base (producto de la monta de otras vacas);
- Mugidos frecuentes, gestos ansiosos, huele la orina de otra vacas (nerviosismo);
- Aumento de la actividad física (inquietud);
- Aumento de la temperatura corporal;
- Disminución del consumo de forraje y de los tiempos de rumia;
- Disminución en la producción de leche; y
- Formación de grupos sexualmente activos entre los animales que se encuentran en celo, con aquellos que están a pocas horas de entrar o salir del mismo.

Entre los métodos conocidos para determinar cuando una vaca está en celo se incluyen:

- **Observación directa** de los animales por un operario para detectar visualmente los síntomas de celo antes enumerados.

- **Uso de retajos**, que son animales que no pueden copular pero mantienen su libido (machos operados o hembras androgenizadas) y por lo tanto cuando montan a una hembra no la sirven pero pueden ser observados por el operario;
- **Dispositivos de detección de monta**, que utilizan pinturas sensores sensibles a la presión colocados sobre la base de la cola de la vaca. Cuando la vaca es montada por otro animal se despinta o el dispositivo cambia de color o proporciona alguna otra indicación del evento que puede ser observada, y
- **Dispositivos de detección de actividad**, que utilizan sensores mecánicos o electrónicos de movimiento colocados en las extremidades (podómetros) o el cuello de la vaca (collares) para medir el aumento de actividad durante el celo. Los podómetros miden la cantidad de pasos realizados por el animal mientras que los collares miden el movimiento general del animal. Cuando la vaca cambia el nivel de actividad física, el dispositivo proporciona alguna indicación del evento que puede ser observada.

Todos estos métodos son ayudas importantes para la detección de celo pero son muy imprecisos y costosos en tiempo. Por lo tanto sería importante desarrollar un sistema que provea este tipo de información a los productores y técnicos en forma relativamente económica y automática. Como se ha descrito, durante las hora que dura el celo o estro, el comportamiento de los animales cambia notablemente en diversos aspectos (monta de un animal, inquietud, tiempos de consumo y rumia más cortos y tendencia a agruparse con otras vacas, entre otros), y la observación de estos síntomas podrían automatizarse. Hasta el momento, para la detección de celo solo se han desarrollado dispositivos y métodos para detectar la monta de los animales o sus cambios de nivel de actividad física. De hecho, la medición del nivel de actividad física a través del análisis del movimiento del cuerpo del animal es una técnica conocida y no invasiva que se puede realizar con sensores comerciales (Sistemas Heatime HR y Heatime Pro de SCR, sistema ALPRO de DeLaval). La tendencia a agruparse que se ve acompañada por intentos de monta entre animales de estos grupos no ha sido considerada en la automatización. Finalmente, uno de los cambios más evidentes del estado de celo se puede registrar en los cambios de la conducta alimentaria del animal. En este sentido, la cantidad de materia ingerida y el tiempo que el animal pasa rumiando disminuyen. Entonces, es posible detectar el estado de celo a partir de un análisis de los cambios en el comportamiento alimentario, lo cual se puede realizar a través del análisis del sonido producido por el animal durante la masticación del forraje, tanto durante la ingestión como durante la rumia. Esta es una técnica conocida y no invasiva que se puede realizar con sensores comerciales y métodos conocidos y adaptarse para su uso en la detección automática de celo. Las únicas implementaciones disponibles de este enfoque consisten en la medición de los cambios de los periodos de rumia, la cual se combina con cambios en el nivel de actividad física. Cuando una vaca aumenta la actividad y disminuye la rumia (asociada a un menor consumo) respecto a los días anteriores, estaría en celo.

En general, los dispositivos para la detección de celo se han concentrado en dispositivos electrónicos que utilizan sensores de movimiento, están montados sobre las extremidades (podómetros) o el cuello del animal (collares), **para medir solo el cambio de actividad ignorando los cambios en los comportamientos alimentarios y grupal de los animales**. Los podómetros miden la cantidad de pasos realizados, no son fáciles de instalar y pueden dar información muy imprecisa. En sistemas de base pastoril los cambios diarios de actividad pueden deberse a otras razones (cambios de parcelas o lotes) y no a la presencia de celo y por lo tanto se indican "*falsos positivos*". Además, las condiciones ambientales hacen que los dispositivos fallen con frecuencia y/o pueden causar molestias y lesiones al animal. Los dispositivos montados en el cuello miden el movimiento general del animal, pero también son sensibles a otras actividades

(comer, oler y jugar), por lo que también frecuentemente proveen información imprecisa.

Hasta el momento se ha utilizado una gran variedad de sensores para detectar actividad. Por ejemplo, un técnica conocida para medir la actividad utiliza una ampolla de cristal que contiene electrodos y una gota de mercurio. Cuando existe movimiento, el mercurio cubre los electrodos y la impedancia entre los electrodos disminuye. Otra metodología utilizada para medir la actividad consiste en una bola de metal que crea un campo electromagnético al cambiar de ubicación. En los últimos años, los sensores de aceleración basados en tecnología MEMS han sido utilizados para medir la actividad de los animales. Sin embargo, la gravedad afecta a estos sensores de manera que la inclinación del sensor se modifica y cambia la lectura con el movimiento, por lo que requieren de procesamiento extra para volver a calibrarlos.

La patente US 4.846.106 describe un interruptor electrónico montado en parte trasera de la vaca que se activa cuando otra vaca la monta.

La patente US 4.247.758 describe un dispositivo que mide e indica el número de movimientos de los animales para la detección de celo.

La patente US 4.618.861 describe un dispositivo que mide la actividad de un animal, y utiliza el movimiento del dispositivo como fuente de energía.

La patente US 5.111.799 describe un interruptor electrónico colocado en la parte trasera de la vaca y se activa cuando otra vaca monta.

La patente US 5.542.431 describe un sistema para determinar la cantidad de veces que un animal es montado a partir de un interruptor ubicado en los cuartos traseros del animal.

La patente US 5.881.673 describe un sistema para determinar la cantidad de veces y el tiempo que un animal es montado a partir de un interruptor ubicado en la zona de la cola del animal.

La patente US 7.083.75 B1 describe un dispositivo que utiliza un interruptor en el área de la cola para detectar la cantidad de veces que un animal es montado.

La patente US 7.137.359 B1 describe un dispositivo que utiliza un sensor de presión para detectar cuando un animal monta a otro.

La patente US 7.230.535 B2 describe un dispositivo que utiliza un interruptor electrónico colocado en la vaca y se activa cuando otra vaca monta. Esta información es utilizada para determinar cuando el animal está en celo y luego es presentada al usuario.

La patente US 7.878.149 B2 describe un dispositivo y un método que utiliza un inclinómetro para medir los movimientos de la cabeza del animal, y su actividad, y luego a través de un análisis estadístico se detecta cuando está en celo.

La patente US 7.927.287 B2 describe un método y un dispositivo para determinar el estado de reproductivo de los animales a partir de la identificación de los eventos de cría (celo, monta, embarazo, parto, entre otros). El dispositivo proporciona de una señal indicativa cuando un animal está en celo utilizando una capa adhesiva que se fija en la protuberancia del animal. Sobre la superficie superior de la capa adhesiva está estratificada una capa indicadora que está cubierta con una capa de revestimiento. La capa de revestimiento está adaptada para su eliminación tras el montaje, exponiendo la capa indicadora y generando una indicación de monta.

La patente US 7.992.521 B2 describe un dispositivo para detectar automáticamente el apareamiento de animales. El dispositivo incluye un detector de intento de apareamiento y un identificador electrónico que se introduce en el cuerpo de la hembra y se acciona por el detector. El identificador incluye un transceptor con una etiqueta electrónica. El transceptor puede leer el identificador electrónico del animal que intenta montar. El dispositivo puede almacenar información

representativa sobre el intento de monta y luego comunicarla a través del transceptor.

La patente US 8.066.179 B2 describe un método y un sistema que permite detectar e identificar cuando un animal monta a otro a partir de la posición de los animales utilizando dispositivos de identificación por radiofrecuencia (RFID).

La patente US 8420398 B2 describe un método para determinar la fase del ciclo estral en el que se encuentra un animal en el momento que se obtiene la muestra biológica. La muestra biológica se combina con al menos un pigmento flavonoide que reacciona con la muestra sobre una superficie hidrófoba para proporcionar una respuesta de color. La fase estro del ciclo estral tiene una respuesta de color correspondiente que se puede distinguir sin asistencia.

La patente US 8.662.021 B2 describe un dispositivo y un método que utiliza un sensor electromecánico ubicado en una de las extremidades para determinar los tiempos que el animal permanece parado o echado.

La patente US 8.949.034 B2 describe un método para determinar cuando un animal está en celo y su estado de salud a partir del análisis su ubicación y movimientos en establos. El método propuesto estima y analiza parámetros característicos del movimiento de los animales como son el porcentaje de tiempo que los animales están caminando y el número de veces los animales se trasladaron una distancia predefinida durante un periodo de tiempo dado, entre otros. El método propuesto supone que se dispone de un sistema de medición y registro de los desplazamientos de los animales en el corral utilizando dispositivos RFID y sistemas de comunicación de banda ultra ancha.

La patente US 8979757 B2 describe un método y un sistema para controlar el estado del ganado, el cual comprende una pluralidad de sensores para detectar los parámetros relevantes del comportamiento de los animales. La información registrada es transmitida inalámbricamente por el dispositivo montado en el animal a un procesador central, el cual determina una el estado del animal la proximidad de un parto, el estado de fertilidad y otros estado de salud del animal.

Las patentes US 20140107434 y US 20150202033 A1 describen un método para la detección de celo que utiliza información de la ubicación de los animales dentro de un establo. Esta información incluye la ubicación de los animales a intervalos regulares durante un periodo de tiempo. Esta información es utilizada para generar indicadores de conductas (tiempo parado, tiempo echado, tiempo de bebida, conteo de montas, entre otras), la cual es utilizada para determinar el estado reproductivo del animal.

La patente US 20140121558 A1 describe un dispositivo o conjunto de dispositivos para medir en forma remota, continua y autónoma parámetros biológicos y localizar los animales. Esto permite la detección temprana de enfermedades y otros estados biológicos, especialmente brotes epidémicos o robo de ganado, facilitando así la reacción temprana de los propietarios y/o autoridades.

La patente US 20160165851 A1 describe un dispositivo fijado al cuello de un animal que utiliza un acelerómetro para producir señales indicativas de los movimiento del cuerpo y la cabeza del animal por periodos de quince minutos. Estas señales son analizadas para detectar los estados de los animales: reposo, alimentación y tres niveles de actividad física del animal. Los datos indicativos de los estados del animal son almacenados en el dispositivo y transmitidos periódicamente a un servidor.

Ninguna de estas referencias sugieren un método para detectar el celo mediante la combinación de la información de los periodos de alimentación, comportamiento grupal, e información sobre el nivel de actividad física del animal. Por lo tanto hay una necesidad de un sistema que permita un

seguimiento más fiable de los ciclos de celo y actividades de cría de ganado.

### ***Resumen de la invención***

La presente invención proporciona un método y un dispositivo no invasivo para la detección de celo en animales mediante la recopilación de información relacionada con la actividad física, la alimentación y el comportamiento grupal del animal. Esta información se combina para reducir el impacto de los hábitos y situaciones anómalas de los animales en el proceso de detección.

El método y dispositivo de la presente invención tiene la ventaja de utilizar y vincular toda la información disponible de los diferentes aspectos del comportamiento del animal de forma no invasiva. Esta capacidad mejora la interpretación de la información obtenida por los sensores, logrando así reducir los errores de estimación. La integración de la información resulta esencial, ya que cada una de las fuentes de información utilizadas por separado tiene sus limitaciones y problemas:

- **Comportamiento físico:** gran parte de los movimientos percibidos del animal se producen durante su alimentación y descanso, de modo que cambios en los lugares de alimentación y descanso generan cambios en la actividad física del animal;
- **Comportamiento ingestivo:** los cambios en la conducta alimentaria también pueden ser provocados por enfermedades o cambios en la dieta; y
- **Comportamiento grupal:** las vacas suelen agruparse para pastorear y rumiar.

Esta invención proporciona un dispositivo no invasivo y un método que comprende las etapas de detección y análisis las de señales indicativas de la actividad física (**comportamiento físico individual**), alimentación (**comportamiento alimentario**) y **comportamiento grupal** de un animal, identificando cuando los comportamientos analizados se desvían, en su conjunto, del comportamiento normal. La identificación de los períodos de alimentación y sus modificaciones, sumados a los cambios en el comportamiento grupal de cada animal, ayudan a minimizar las falsas interpretaciones de los cambios de la actividad física fuera de la conducta normal, a fin de obtener identificaciones más precisas y exactas del estado de celo. La detección se realiza mediante el análisis de los comportamientos alimentario, grupal y actividad física del animal utilizando clasificadores basados en herramientas de inteligencia computacional y aprendizaje maquina (lógica borrosa, árboles de decisión, redes neuronales, entre otras) o estadísticas.

Esta invención proporciona además un dispositivo que comprende sensores para detectar la actividad física, la actividad alimentaria y el comportamiento grupal de un animal. Un sistema embebido basado en microprocesadores, dispositivos lógicos programables (FPGA) o circuitos integrados de aplicación específica (ASIC), registra y analiza los datos de movimiento, posición, y sonido relevados. Cada una de estas señales es procesada de manera independiente para acondicionarlas y extraer la información indicativa correspondiente a cada uno de los aspectos de la conducta del animal considerados. El dispositivo puede, al menos parcialmente, estar contenido dentro de una caja estanca que se colocará preferentemente, de acuerdo con una realización de la invención, donde se puedan detectar la inclinación y el movimiento de la cabeza del animal, el sonido generado al alimentarse y la posición del animal. En otras formas de realización de la invención, el dispositivo puede ser montado en el cuello del animal usando un cinturón. Un medio indicador que muestre cuando el animal está en celo puede estar acoplado a la caja. El proceso de detección puede realizarse su totalidad en el sistema embebido y una notificación puede ser enviada cuando el animal se identifica como en celo. Otra alternativa consiste en que sólo una parte del proceso pueda llevarse a cabo en el sistema embebido y la información registrada y procesada pueden transmitirse a un servidor remoto para detectar el estado del animal.

El análisis del **comportamiento físico individual** de los animales puede basarse en medidas estadísticas de los datos de movimiento registrados comparados con una base de datos de referencia. Otra posibilidad consiste en la implementación de clasificadores basados en herramientas de inteligencia computacional y aprendizaje maquina (redes neuronales, árboles de decisión, etc.) que clasifiquen los datos registrados, o parámetros característicos de la señal, para determinar el estado del animal. Los sensores de movimiento preferidos son las unidades inerciales de medición (IMU) basadas en tecnología MEMS y el GPS. Las IMUs incluyen en un dispositivo acelerómetros, giróscopos y compases magnéticos que permiten corregir, a partir de procesamiento de señales, los errores de los sensores y compensar los efectos de la gravedad. Analizando y combinando las señales corregidas de la IMU se puede extraer información de los movimientos que realiza el animal. Esta información permite definir, por ejemplo, el nivel de los movimientos del animal, los momentos en que realiza las montas o identificar cualquier otro tipo de acciones que sean indicativas del celo. El nivel de la actividad física puede obtenerse a partir de la energía o del valor medio cuadrático de la señal de movimiento, corregido sobre diferentes intervalos de tiempo. Los momentos de montas pueden identificarse a partir de la detección de cambios abruptos y de corta duración de la aceleración en la dirección vertical. Para detectar otro tipo de acciones se puede utilizar un sistema reconocedor que combine herramientas de procesamiento de señales y aprendizaje maquina, el cual puede funcionar en dos etapas. En la primera etapa se transforma la señal corregida de la IMU en un conjunto de características predefinidas (información frecuencial, atributos estadísticos, energía, etc.). En la segunda etapa, un clasificador utiliza un conjunto de características y asigna a su salida el estado del animal que corresponda.

Para medir el movimiento del animal, los sensores pueden estar montados en el cuello o la cabeza del animal. Si están ubicados en el cuello del animal, los sensores simplificarán la tarea de detección, ya que solo registran el movimiento del cuerpo del animal. Sin embargo, si los sensores están ubicados en la cabeza, estos medirán el movimiento combinado de la cabeza y el cuerpo. En este último caso, se medirá la inclinación y los movimientos de la cabeza animal en relación con el suelo, permitiendo detectar los periodos de alimentación y mejorar así la estimación de la actividad alimentaria ya que cuando el animal come, la cabeza del animal baja hacia el suelo y moverse, mientras que por lo general el animal tiende a mantener su cabeza más alta cuando está activo y/o rumiando. En celo el porcentaje de tiempo que el cuello del animal está abajo (comiendo) es significativamente más pequeño en comparación con el de los periodos normales ya que el animal consume menos pastura durante el celo. Una combinación de alto nivel de movimiento de la cabeza y altos valores de inclinación de la cabeza puede indicar con una alta probabilidad que el animal está alimentándose.

Además de aumentar su actividad física general, los animales en celo tratan de montar a otros animales. Además, si estos se dejan montar, es una comprobación de que el animal está en celo. En general el proceso de monta se puede detectar a través de dos sensores: i) a través de sensores de movimiento (acelerómetros, inclinómetros, GPS, etc.) que detectan el momento del salto y ii) a través de dispositivos que detecten la proximidad de dos animales, como las redes de comunicación de campo corto (Near Field Communication - NFC) y los dispositivos de identificación de radiofrecuencia (Radio Frequency Identification Device - RFID). Estos dos dispositivos tienen en común que la distancia de comunicación entre emisor y receptor puede controlarse, garantizando una distancia máxima de comunicación. Otra forma de detectar la proximidad de dos animales es compartiendo vía red de comunicación y comparando la posición registrada por un GPS de cada uno de ellos.

En el caso de utilizar acelerómetros, se busca una variación de gran magnitud en la señal de



aceleración vertical dado que el animal que monta a otro debe elevarse con un salto y al caer choca con el cuerpo del animal que está debajo. Este movimiento veloz hacia arriba y hacia abajo se manifiesta distintivamente en la amplitud de la aceleración porque en un corto período de tiempo se producen variaciones que no están presentes cuando el animal desarrolla el resto de sus actividades. También podría entrenarse un reconocedor para detectar estos eventos a partir de las señales medidas de aceleración.

El análisis del **comportamiento alimentario de los animales** puede basarse en el análisis de los sonidos que normalmente produce el animal cuando se alimenta. Cuando el animal realiza alguna actividad alimentaria (pastoreo o rumia) su mandíbula genera sonidos que pueden ser registrados a través de un sensor y luego analizados para determinar qué actividad alimentaria realiza. La cabeza y el cuello del animal son ejemplos de áreas adecuadas para el registro de los sonidos. El sensor de sonido preferido son los micrófonos electret direccionales, los cuales son capaces de registrar sonidos muy débiles en la dirección de preferencia (frente del micrófono), sin embargo cualquier otro tipo de micrófono puede ser utilizado. Como la señal es débil, debe ser acondicionada (amplificada y filtrada) a través de circuitos electrónicos analógicos específicos y luego digitalizada. Se pueden utilizar dos configuraciones: *i*) un solo micrófono si la aislación acústica es buena de modo que la señal registrada tenga una buena relación señal/ruido, ó *ii*) dos micrófonos para poder realizar una cancelación activa del ruido ambiente en condiciones más desfavorables. En ambos casos, la señal es procesada por un algoritmo de limpieza que atenúa los efectos del ruido ambiente y otras señales no deseadas. La señal sonora resultante es analizada para determinar cuando el animal está ingiriendo alimento o rumiando. Los sonidos que se producen durante la ingesta de alimento son más intensos y cortos que cuando el animal está rumiando, los cuales son más suaves y regulares con frecuencias de repetición diferentes. La razón de esta diferencia estriba en los componentes que actúan en cada uno de los mecanismos, además de la consistencia y el contenido de agua que tiene la materia procesada y la disponibilidad de la misma. De modo que utilizando herramientas de procesamiento de señales e inteligencia computacional, es posible determinar cuando el animal se está alimentando o rumiando. De modo que una vez identificados los períodos y parámetros de las actividades de ingesta y rumia se pueden acumular los parámetros detectados, y utilizando métodos estadísticos o herramientas de inteligencia computacional se pueden determinar parámetros de referencia para caracterizar las actividades alimentarias normales.

La actividad alimentaria del animal también se puede evaluar de manera complementaria mediante el análisis de la posición de la cabeza respecto al suelo y sus movimientos. Las señales generadas por los acelerómetros incluyen información de la posición de la cabeza, a partir de la de la orientación del peso de la cabeza (fuerza de gravedad), así como los movimientos que realiza. Una pequeña distancia entre la cabeza y el suelo, sumado a un aumento de movimiento pueden indicar que el animal está comiendo. La medición de la distancia se puede realizar utilizando acelerómetros y giróscopos basados en tecnología MEMS. Esta combinación de sensores permite medir tanto el ángulo de inclinación de la cabeza como su movimiento. De este modo es esperable que, cuando el sensor de inclinación del cuello mida un ángulo grande (la cabeza del animal cerca del suelo) y la intensidad de los movimientos de la cabeza sea alta, es probable que el animal se esté alimentando. Por el contrario, si el ángulo es pequeño (la cabeza del animal lejos del suelo) y la intensidad de los movimientos de la cabeza es baja, es probable que el animal esté rumiando. De esta forma, el seguimiento estadístico de las señales de estos sensores puede utilizarse para evaluar el estado reproductivo del animal

El análisis del **comportamiento grupal de los animales** puede basarse en el uso de técnicas de clustering. Estas técnicas permiten encontrar los grupos de animales con comportamientos

similares combinado con otros datos, como las distancias medias relativas entre los animales a lo largo del tiempo, proporcionar una estimación de grupos candidatos a estar en celo y factibles de ser estudiados utilizando información complementaria. Cuando un animal está próximo o en celo, tiende a agruparse con animales que están o acaban de terminar el celo. Este comportamiento puede ser monitoreado a través de un dispositivo que determine los animales que se hallan próximos y comparando la información obtenida con una base de datos de los animales en celo del rodeo. Luego se analiza estadísticamente la composición del grupo de animales comparando con la información adquirida con una base de datos del estado de los animales. La base de datos se actualiza de forma periódica para seguir los cambios de estado de los animales a lo largo del tiempo.

Cada animal del rodeo es identificado a través de un código único, el cual se halla almacenado en los dispositivos. Una forma de detectar la proximidad de los animales es a través del uso de redes NFC. Estos dispositivos permiten intercambiar información controlando la distancia de comunicación entre emisor y receptor. De esta manera se pueden explorar los animales que se hallan a determinada distancia de un animal dado. Otra forma de detectar la proximidad entre animales es compartiendo, vía red de comunicación, el código de identificación y la posición de los animales. Una vez detectados los animales próximos al animal bajo estudio, se compara esta información con una base de datos de los animales en celo en ese momento. A partir de un análisis estadístico de esta información, es posible determinar la probabilidad de que el animal bajo estudio se encuentre en celo dada la información del comportamiento grupal del mismo. Además, este bloque permite detectar cuando un macho específico monta a una vaca y así poder realizar un seguimiento de la genética del rodeo.

Las informaciones correspondiente a cada uno de los comportamientos analizados en cada animal (alimentario, actividad física y grupal) se integran a lo largo del tiempo con cualquier otra información relevante (fecha estimada de celo, producción láctea, etc.) para determinar si un animal está en celo o no. El análisis puede basarse en el uso de técnicas de aprendizaje maquina (máquinas de soporte vectorial, redes neuronales, árboles de decisión, etc.) o análisis estadísticos que permitan determinar cuándo el comportamiento del animal se aleja del comportamiento normal y se aproxima a su comportamiento durante el celo. El uso de estas fuentes de información permite distinguir aquellos cambios en solo uno de los aspectos del comportamiento del animal, disminuyendo el número de falsas alarmas producidos por problemas alimentarios, cambios de dietas y cualquier otra situación anómala que pueda afectar la caracterización del celo en un animal dado. Una vez determinado el estado del animal se lo comunica al sistema.

### ***Breve descripción de las figuras***

La presente invención se entenderá y apreciará mejor a partir de la siguiente descripción detallada tomada en conjunto con las siguientes figuras:

Figura 1. Diagrama de bloques que ilustra una posible realización del dispositivo de acuerdo con la invención mostrando sus bloques fundamentales;

Figura 2. Ilustración esquemática de la ubicación del dispositivo, los sensores y la antena;

Figura 3. Diagrama de bloques que ilustra una posible realización del bloque de análisis de comportamiento alimentario;

Figura 4. Ilustración de la salida eléctrica de la unidad de sensor de sonido de 30 segundos durante a) pastoreo y b) rumia;

Figura 5. Ilustración de diferentes posiciones de la cabeza de un animal: a) cuando está en grupo,

b) cuando rumia y c) cuando se alimenta;

Figura 6. Diagrama de bloques que ilustra una posible realización del clasificador de actividades alimentarias;

Figura 7. Diagrama de bloques que ilustra una posible realización del bloque de análisis de comportamiento físico individual;

Figura 8. Ilustración de la salida eléctrica de la unidad de medición inercial para los movimientos de un animal durante comportamiento a) normal y b) celo;

Figura 9. Diagrama de bloques que ilustra una posible realización del clasificador de actividad física individual y detección de monta;

Figura 10. Diagrama de bloques que ilustra una posible realización del bloque de análisis de comportamiento grupal;

Figura 11. Ilustración del proceso de monta y la comunicación entre los dispositivos utilizando una red de comunicaciones de campo corto (NFC);

Figura 12. Diagrama de bloques que ilustra la realización del bloque de análisis de comportamiento del animal, almacenamiento y comunicaciones;

Figura 13. Diagrama de bloques que ilustra una posible realización del bloque de alimentación;

Figura 14. Diagrama de bloques que ilustra la implementación concurrente del dispositivo utilizando cuatro procesadores.

Se apreciará que por simplicidad y claridad de la ilustración, los elementos mostrados en las figuras no han sido necesariamente dibujados a escala. Por ejemplo, para mayor claridad las dimensiones de algunos de los elementos pueden estar exageradas con relación a otros elementos. Además, cuando se considera apropiado, los números de referencia pueden ser repetidos entre las figuras para indicar elementos correspondientes o análogos.

### ***Descripción detallada de la invención***

La presente invención utiliza la información indicativa de la actividad física (movimientos del animal) en combinación con información indicativa de la actividad alimentaria (sonidos y posición de la cabeza) e información indicativa de su comportamiento grupal (estructura del grupo) para determinar si un animal está en celo o no. La integración de estas tres fuentes de información permiten mejorar la precisión y exactitud de la detección de celo al interpretar correctamente los cambios de comportamiento del animal.

Esta invención incluye un dispositivo que comprende los sensores para registrar los movimientos del animal, su posición, los sonidos producidos durante la alimentación e identificar los animales que se hallen próximo a él. Un sistema embebido basado en microprocesadores, dispositivos lógicos programables (FPGA) o circuito integrado de aplicación específica (ASIC), registra, procesa y analiza todas estas señales de manera independiente para extraer la información indicativa de cada uno de estos aspectos de la conducta del animal. El proceso de detección de celo puede realizarse en su totalidad en el sistema embebido y la información resultante ser enviada a una computadora periódicamente. Otra alternativa consiste en que sólo una parte del proceso se lleve a cabo en el dispositivo (por ejemplo el registro y acondicionamiento) y la información obtenida se transmite a una computadora para su análisis. El dispositivo puede estar contenido parcialmente dentro de una caja estanca que se ubicará preferentemente en la cabeza o el cuello del animal, de acuerdo con una realización de la invención. Estos lugares permiten detectar fácilmente los movimientos y posición del animal, sumado a que cuando se alimenta

(rumia o pastoreo), se generan sonidos que pueden ser relevados a través de un sensor en la cabeza del animal y luego analizados para determinar si el animal está en celo o no. Los sensores de sonido preferidos son los micrófonos electret direccionales, los cuales son capaces de registrar sonidos muy débiles en la dirección de preferencia y su respuesta en frecuencia en la banda de interés es plana. Como la señal es débil, debe ser acondicionada (amplificada y filtrada) a través de circuitos electrónicos analógicos específicos y luego digitalizada. La señal de audio digitalizada es procesada previamente por un algoritmo de limpieza, el cual atenúa el efecto del ruido ambiente y otras señales no deseadas. Los sensores de movimiento preferido son las unidades de medición inercial (IMU), las cuales son capaces de registrar las aceleraciones traslacionales, las velocidades angulares y la posición con respecto al norte magnético de la tierra. Estas señales son procesadas por algoritmos de estimación y filtrado para remover los efectos de los ruidos y errores (offsets y bias). Finalmente, el sensor de posición preferido es un GPS, el cual puede incorporar alguna señal de corrección para mejorar la precisión de los datos medidos.

La Figura 1 presenta el diagrama de bloques de un dispositivo de detección de celo construido y funcionando de acuerdo con las realizaciones de la presente invención. Incluye seis bloques funcionales:

1. **Adquisición y análisis de comportamiento alimentario:** este bloque es el encargado de registrar los sonidos y movimientos de la cabeza producidos por el animal al alimentarse para determinar y cuantificar la actividad alimentaria del animal. Las señales registradas son procesadas y acondicionadas para remover las perturbaciones y ruidos que se puedan haber registrado junto a los sonidos producidos por el animal al alimentarse. La señal del movimiento de la cabeza es procesada específicamente para remover los errores sistemáticos de los sensores y los ruidos. Luego, los sonidos son analizados, e integrados con los movimientos y posición de la cabeza, para detectar y cuantificar en tiempo real las actividades alimentarias (pastoreo o rumia) realizadas por el animal.
2. **Adquisición y análisis de comportamiento individual:** este bloque es el encargado de registrar los movimientos del animal a lo largo del tiempo para determinar y cuantificar la actividad física realizada por el animal. Las señales registradas son las aceleraciones traslacionales, las velocidades de rotación y la orientación de la cabeza con respecto al norte magnético. Luego, estas señales son procesadas y acondicionadas para remover los errores sistemáticos y ruidos que se puedan haber registrado junto a ellas. La señal del movimiento del animal es analizada para detectar y cuantificar en tiempo real el nivel de actividad física desarrollada por el animal y detectar momentos de monta realizados por el animal. El análisis de la señal de movimiento también permite obtener un historial de las actividades físicas desarrolladas por el animal, lo que posibilita la comparación de historiales de distintos días para detectar cambios de comportamiento.
3. **Adquisición y análisis de comportamiento grupal:** este bloque es el encargado de registrar la composición y estructura del grupo al cual pertenece el animal a lo largo del tiempo. Las señales registradas son la posición del animal en el campo y los animales que se encuentran próximos al mismo. Estas señales son procesadas y acondicionadas para remover los errores y ruidos que se puedan haber registrado junto a ellas, y son analizadas para detectar y cuantificar en tiempo real la cantidad de animales en celo que hay dentro el grupo en el que está animal.
4. **Análisis de comportamiento reproductivo:** este bloque es el encargado de integrar y analizar la información generada por los bloques anteriores con cualquier otra información disponible para detectar si el animal está en celo o no.

5. **Almacenamiento y comunicaciones:** este bloque es el encargado de organizar toda la información generada por el dispositivo a partir de la información registrada. Se encarga de organizar la información en bloques autocontenidos que proveen información adicional de utilidad (por ejemplo el lugar y hora donde se registraron los datos, entre otros). Además, es responsable de intercambiar información a través un sistema de comunicación inalámbrica. Esto permite comunicar al sistema de gestión la información del animal en tiempo real.
6. **Administración de energía:** este bloque es el encargado de administrar la energía y la disponibilidad de las baterías del dispositivo para maximizar su tiempo de operación. Esta tarea se realiza monitoreando la carga de la batería, la energía provista por el sistema de recolección de energía y deshabilitando funciones del dispositivo, de acuerdo a su prioridad.

Estos bloques se describirán en detalle a continuación.

La Figura 3 ilustra la realización del módulo de análisis de actividad alimentaria. La señal sonora registrada por el micrófono sigue dos caminos diferentes: *i*) uno para detectar y clasificar los eventos y *ii*) el otro para calcular los parámetros característicos de la actividad. Cada uno de estos caminos recibe diferentes acondicionamientos antes de su procesamiento. En ambos caminos la señal es filtrada y amplificada para reducir los efectos del ruido y las perturbaciones, y luego un conversor analógico/digital digitaliza la señal. La señal utilizada para detectar los eventos es acondicionada con un amplificador con control automático de ganancia (AGC) para obtener una señal con buena relación señal/ruido y sin distorsiones. Esto mejora la tasa de reconocimiento y clasificación, en relación al caso en que se la procese con un amplificador con ganancia fija. Sin embargo, esto dificulta el cálculo de los parámetros característicos de la actividad alimentaria que están relacionados con la energía de la señal, ya que la señal original es modificada por una ganancia variante en el tiempo y desconocida. La señal utilizada para calcular estos parámetros es acondicionada por un amplificador de ganancia fija y conocida, de modo que puede determinarse la relación entre la señal original y la procesada, y de ese modo calcular la energía contenida en la señal original. Para captar los sonidos producidos por el animal se utiliza un micrófono electret direccional sujeto contra la frente del animal en U2 (Figura 2).

Los movimientos de la cabeza son registrados a través de una unidad de medición inercial (IMU), la cual registra las aceleraciones traslacionales y la velocidades angulares asociadas al movimiento de la cabeza. La posición angular puede determinarse con los acelerómetros a partir de la estimación del vector de la gravedad y el uso de funciones trigonométricas. Sin embargo, los acelerómetros miden todas las fuerzas que están actuando sobre el objeto, por lo que las mediciones pueden estar contaminadas con perturbaciones de corta duración, haciendo que los datos de los acelerómetros sólo sean fiables en el largo plazo. Por otro lado, la posición angular también puede determinarse con los giróscopos, los cuales miden las velocidades angulares. Sin embargo, estos sensores tienen derivas (no retornan a cero cuando el objeto retorna a su posición original) que introducen errores sistemáticos de largo plazo. Esto hace que los datos de los giróscopos sólo sean fiables en el corto plazo. Para obtener mediciones precisas y confiables es necesario realizar correcciones a ambas mediciones combinadas. Con las señales de aceleración y velocidades angulares corregidas el bloque de estimación calcula los movimientos y la posición de la cabeza. Para registrar los movimientos de la cabeza se utiliza una IMU localizada en el dispositivo en U1 (Figura 2).

La Figura 4 presenta la salida del micrófono durante 30 s en pastoreo (Figura 4.a) y rumia (Figura 4.b). Como se puede apreciar en la Figura 4.a, los sonidos que se producen durante el pastoreo se caracterizan por ser de corta duración y sin un ritmo tan definido. Por otro lado, el sonido producido durante la rumia se caracteriza por su morfología temporal, menor energía relativa y

una periodicidad más marcada a dos niveles: el nivel de las masticaciones durante la rumia (con una frecuencia alrededor de 1 Hz) y el nivel de las regurgitaciones (que se repite aproximadamente cada un minuto). El animal rumia echado o parado y hay en promedio de 15 a 20 períodos diarios donde se regurgitan de 300 a 400 porciones de alimento con un promedio de 50 movimientos masticatorios por porción.

El primer paso para analizar la actividad alimentaria es identificar y caracterizar los eventos masticatorios: mordida, mordida-masticación y masticación. Cada una de estas acciones producen sonidos cuyo contenido frecuencial estará dado por el material procesado, mientras que la duración y amplitud del sonido estarán determinados por el movimiento de la mandíbula. Entonces, el sonido producido por una mordida se caracteriza por una gran amplitud y una corta duración, característico de un rasguído. En cambio, el sonido producido por una masticación se caracteriza por una pequeña amplitud y una duración más larga, característico de un chirrido. El movimiento combinado masticación-mordida es la combinación de ambos. Finalmente, se considera ruido a un sonido que no satisface ninguna de las características antes descritas y un silencio es la ausencia de sonidos durante el periodo de un evento promedio.

El comportamiento alimentario del animal también puede evaluarse a través de la posición de la cabeza respecto al suelo y sus movimientos. Una distancia pequeña (ángulo grande) y un aumento de la cantidad de movimientos locales de pequeña amplitud pueden indicar que el animal está comiendo (Figura 5.c), mientras que una distancia grande (ángulo pequeño) y pocos movimientos locales pueden indicar que el animal está rumiando (Figura 5.b). Para medir la posición y movimientos de la cabeza se utilizan los acelerómetros de la IMU. Estos sensores permiten medir tanto el ángulo de inclinación como los movimientos de la cabeza. En la Figura 5 se muestran ilustraciones de varias posiciones de la cabeza de un animal en diversas situaciones. Cuando el animal no está comiendo, la inclinación puede ser como se ilustra en las Figuras 5.a y 5.b. Cuando el animal ingiere alimento, la inclinación aumenta ya que la cabeza baja hacia el suelo como se ilustra en la Figura 5.c. Cuando el animal está rumiando, la inclinación de cabeza puede ser como se ilustra en la Figura 5.b. Un método popular para interpretar los datos de inclinación consiste en el uso de umbrales (fijos o adaptativos) en el ángulo de inclinación de la cabeza. Cuando la señal de inclinación registrada cruza este umbral, significa que la señal de movimiento debe ser interpretada como la representación de un período de alimentación. El umbral utilizado para la clasificación puede ser fijo o actualizarse de manera automática con la información recogida en una ventana de tiempo dado, de manera que los eventos masticatorios que ocurren durante la actividad de alimentación pueden identificarse a partir del análisis de las características temporales más relevantes de los sonidos que produce un animal cuando se alimenta (forma de la envolvente, amplitud máxima, duración, periodicidad, secuencia y simetría, entre otros) combinadas con el análisis de los patrones de movimientos de la cabeza y su ángulo con respecto al piso. Entonces, un clasificador procesa las características temporales en conjunto con patrones de movimiento y ángulos de la cabeza para identificar el evento. Los eventos identificados son: mordida, masticación-mordida, masticación y silencio.

La Figura 6 muestra la realización del bloque de análisis de actividades alimentarias, el cual está constituido por tres subbloques: i) uno para la extracción de características, ii) uno de clasificación y cálculo de parámetros de eventos masticatorios, iii) uno de clasificación y cálculo de parámetros de actividades alimentarias. El extractor de características procesa la señal acústica salida del AGC y la señal de la posición y movimientos de la cabeza para extraer las características temporales más relevantes de la señal. En una primera etapa se detecta la envolvente de la señal de audio registrado, la cual es submuestreada. En esta implementación se calculan cuatro características temporales: amplitud máxima, duración, forma (signo de la derivada) y simetría. Pero pueden



calcularse otras. La amplitud máxima se detecta directamente desde la señal sonora, en una ventana temporal con el tamaño de un evento masticación-mordida, ya que es el evento más largo. Las restantes características son determinadas a partir de la envolvente de la señal.

Estas características de las señales sonoras y los movimientos y posición de la cabeza son analizados en conjunto por el clasificador de eventos. Este bloque es el encargado de determinar si ocurrió o no un evento masticatorio y que tipo de evento (masticación, masticación-mordida, mordida) ocurrió. Para realizar esta tarea se pueden utilizar herramientas de inteligencia computacional como las redes neuronales ó los árboles de decisión, entre otros, entrenados a tal fin. En esta implementación se utiliza un perceptrón multicapa que provee la etiqueta del evento. Finalmente, el bloque de cálculo de parámetros de eventos utiliza la información de la etiqueta del evento en combinación con la señal digitalizada sin AGC para calcular la energía de los eventos. Los parámetros característicos restantes (duración, cantidad, amplitud máxima, etc.) son calculados a partir de las características generadas por el subbloque de extracción de características.

Una vez identificados los eventos individuales se procede a identificar y cuantificar las actividades realizadas por el animal en el período considerado a través del clasificador de actividades alimentarias. La información de los eventos (etiqueta y parámetros) es utilizada por el clasificador de actividades, el cual implementa un modelo estadístico basado en árboles de decisión para evaluar la secuencia de eventos identificados (Figura 6). Las actividades identificadas son: pastoreo, rumia y ninguna. Por ejemplo, si se identifica una secuencia de eventos de rumia, con algún silencio o ruido intercalado entre ellos, es altamente probable que el animal estuviera rumiando todo el tiempo, de modo que las etiquetas de los eventos intercalados son corregidas. Una vez determinado el tipo de actividad del bloque de datos se procede a calcular los parámetros característicos del grupo de eventos (tiempo de rumia, tiempo de pastoreo, cantidad de eventos, energía de pastoreo, entre otros) y luego determina la cantidad de materia seca ingerida a partir de un modelo de consumo.

La Figura 7 ilustra la realización del módulo de análisis de comportamiento físico individual. El módulo está compuesto por una unidad de medición inercial (IMU) que mide los movimientos de la cabeza y un GPS que mide la posición absoluta del animal. Estos sensores están ubicados en la cabeza de animal en U1 (Figura 2) o en el cuello. En general, estos sensores tienen errores sistemáticos (bias, drifts, etc.) y aleatorios significativos que dificultan la correcta estimación de los movimientos y posición, por lo que es necesario combinar la información de estos sensores para corregir los errores. El algoritmo utilizado para obtener las señales de movimiento y posición corregidas es un estimador de mínimos cuadrados recursivo. Estas señales corregidas pueden analizarse directamente o utilizarse como entrada para otro bloque de procesamiento.

Durante su actividad diaria, el animal pasa alrededor de trece horas echado descansando y rumiando, cinco horas rumiando y pastoreando parado, tres horas socializando y una hora bebiendo. Estas actividades están combinadas con la actividad corporal y desplazamientos para buscar comida o beber. Las hembras en celo y las que están próximas a él, presentan un mayor dinamismo (caminan al menos un kilómetro por hora), consumen menos alimento (reducción en el pastoreo y rumia) y permanecen más tiempo paradas. La Figura 8 presenta las mediciones del módulo de la aceleración durante un período de actividad normal (Figura 8.a) y un período en el que el animal está en celo (Figura 8.b). Como se puede apreciar en la Figura 8.a, la aceleración durante la actividad normal es de menor intensidad. En cambio, la aceleración durante el celo tiende a ser de mayor intensidad, con picos muy elevados. De todos modos pueden existir períodos de actividad más intensa aún cuando el animal no está en celo, y de allí radica la dificultad para identificar los períodos de celo correctamente solo utilizando la información de

actividad física.

La Figura 9 muestra la realización del bloque de análisis de comportamiento físico individual, el cual está constituido por cinco subbloques: i) estimación y filtrado, ii) segmentación, iii) extracción de características, iv) detección de monta y v) clasificación de actividad física. El bloque de estimación y filtrado es el encargado de estimar la posición del cuerpo, sus movimientos y los movimientos de la cabeza a partir de la información provista por la IMU y el GPS. A partir de un algoritmo recursivo de estimación basado en métodos de mínimos cuadrados, este bloque estima la información utilizada por los algoritmos de detección y clasificación (posición del cuerpo, movimientos del cuerpo y movimientos de la cabeza) minimizando el efecto de los ruidos y errores sistemáticos. Los movimientos del cuerpo se manifiestan, en conjunto con los movimientos de la cabeza, en la señal de aceleración corregida resultante del bloque de estimación y filtrado. Analizando los cambios de posición del centroide geométrico de las posiciones de la cabeza del animal (calculados durante periodos fijos de tiempo) en el plano x-y es posible estimar los movimientos del cuerpo del animal a lo largo del tiempo. Cuando se dispone de mediciones de GPS, es posible corregir esta información. El bloque de segmentación se ocupa de dividir la señal de aceleración en segmentos de unos pocos segundos para entregarle al bloque de extracción de características la señal que debe procesar. Esta división temporal de la señal permite tener una descripción cuasi continua de la actividad física que desarrolla el animal ya que el análisis de cada segmento de la señal de movimiento permite identificar la actividad física desarrollada en ese intervalo. El bloque de extracción de características procesa la señal segmentada para extraer atributos temporales y frecuenciales que rescaten información discriminativa de ella. Como salida del bloque se obtiene un vector de características unidimensional que es la concatenación de todos los atributos obtenidos del segmento analizado. La detección de monta es una característica más, que se adosa al vector de características, aún cuando se ilustre separada en el diagrama de la Figura 9. El bloque de clasificación de actividades está compuesto, en esta implementación, por un clasificador estático (por ejemplo un árbol de decisión, una red neuronal estática o una máquina de soporte vectorial) que es entrenado para reconocer la actividad a la que corresponde cada vector de características que se le pueda muestrear a su entrada. Para realizar dicha tarea, el clasificador, en su fase de entrenamiento, construye modelos de cada actividad física a reconocer, descubriendo similitudes entre los vectores de características de una misma actividad, y estableciendo una relación general entre características y actividades. Esta relación queda fija una vez finalizada la etapa de entrenamiento y es aplicada cada vez que el sistema le presenta un vector de características a este bloque.

Las montas entre animales pueden detectarse analizando la componente vertical de la señal de aceleración, dado que manifiesta el movimiento de ascenso y descenso característico de esta acción (Figura 11). Este proceso puede establecer un umbral que sólo pueda ser superado cuando la amplitud de la variación de la aceleración vertical sea provocado por un salto de monta y luego analizar las características de un periodo de la señal centrado en el posible evento. Este análisis se enfoca en la morfología de la envolvente de la señal de aceleración vertical.

Las características de las señales de movimientos, la posición del animal y la ocurrencia o no de un evento de monta son analizados en conjunto por el clasificador de actividad. Este bloque es el encargado de determinar qué tipo de actividad está realizando el animal y si sus parámetros se desvían del comportamiento normal esperado. Para realizar esta tarea se pueden utilizar herramientas de inteligencia computacional como son las redes neuronales ó los árboles de decisión, entre otros, entrenados a tal fin. En esta implementación se utiliza una red neuronal multicapa que provee la etiqueta de la actividad realizada. Las actividades identificadas son: Parada, Echada, Caminando, Corriendo y Quieta. Una vez determinado el tipo de actividad del



bloque de datos se procede a calcular los parámetros característicos del grupo de eventos.

Si se desea obtener información general del comportamiento a lo largo de un día, puede realizarse un análisis directo de estas señales para obtener, por ejemplo, un historial de cómo fue variando el nivel de actividad del animal o cuántas montas realizó o recibió y en qué momentos. El historial de actividad puede obtenerse al calcular por intervalos de tiempo fijo la energía o la desviación estándar promedio de la señal de aceleración en dicho intervalo. De esta manera no sólo se obtiene el nivel de actividad sino que también se conoce como fue variando la actividad a lo largo del día. Finalmente, el módulo de análisis de actividad provee información relevante a otros módulos del sistema: i) la hora universal, ii) los movimientos de la cabeza y iii) la ocurrencia de eventos de monta. La hora universal se obtiene a partir de la información del GPS, mientras que los movimientos de la cabeza y los eventos de monta son obtenidos a partir de las mediciones de la IMU.

La Figura 10 ilustra la realización del módulo de comportamiento grupal. El módulo está compuesto por una red de comunicaciones de campo corto (NFC) que registra e identifica los animales que se encuentran dentro de una distancia determinada (aproximadamente tres metros) del animal bajo estudio a partir de la ubicación de las antenas de transmisión en el bozal en U3 (Figura 2) y la antena de recepción en el dispositivo en U1 (Figura 2). Utiliza el tiempo medido por el bloque de análisis de comportamiento para realizar un seguimiento temporal de la composición del grupo y así definir una base de datos que describe la composición de grupo. Esta base de datos generada en tiempo real es comparada con una base de datos de los animales en celo del rodeo, la cual es actualizada periódicamente. Esta comparación permite medir la composición del grupo. El grupo puede ser caracterizado a partir de medidas estadísticas del grupo que representan la información relevante del grupo y/o de herramientas de inteligencia computacional. A partir del uso de clasificadores basados en herramientas de aprendizaje maquinal (por ejemplo sistemas borrosos, árboles de decisión, etc.), la información de la composición del grupo y una base de datos de las hembras en celo y los machos disponibles, se puede clasificar el comportamiento reproductivo del grupo al que pertenece un el animal bajo estudio. Esta información es comunicada al módulo de análisis de estado del animal. En el caso de haber detectado un intento de monta, el dispositivo se comunica con la vaca más próxima, que se supone que es la vaca que ha sido montada, para intercambiar los códigos de identificación del animal y así actualizar la información acerca del intento de monta en los dos animales involucrados (Figura 11): el que monta y el que se deja montar.

La Figura 12 ilustra la construcción del bloque de análisis de comportamiento, almacenamiento y comunicaciones. Este bloque es el encargado de determinar el estado del animal y organizar la información generada por los tres bloques de análisis, y otros sensores, para su almacenamiento y transmisión. Este bloque recibe la información de comportamiento alimentario, comportamiento físico individual y comportamiento grupal, procesandola para determinar el estado reproductivo del animal. Para ello, el bloque utiliza herramientas de inteligencia computacional que le permiten integrar la información independiente generada por cada módulo y obtener un resultado más preciso. Además, este bloque organiza la información en un paquete de datos autocontenido al cual se le incorpora un código de detección/corrección de errores sin necesidad de retransmisión de la información. El paquete incluye información de la posición geográfica donde se obtuvo la información, lo cual agrega un elemento de trazabilidad a la información generada por el dispositivo. El paquete de información es almacenado en una memoria no volátil y eventualmente transmitido a una computadora personal para su posterior procesamiento. La transmisión de la información se realizará a través de una interfaz wi-fi de baja potencia. Estas comunicaciones serán periódicas durante el día y la información generada durante la noche se almacenará y será

transmitida a una hora predeterminada para ahorrar energía.

La Figura 13 ilustra la construcción del bloque de administración de energía. Este bloque es el encargado de generar las diferentes líneas de alimentación (analógica y digital) utilizadas por el registrador a partir de una única fuente de energía (baterías) y administrar la energía disponible para maximizar el tiempo de operación del registrador. Esta tarea se realiza de dos maneras complementarias: *i*) asegurando la máxima carga posible de la batería a partir de la energía provista por un panel solar, o cualquier otro dispositivo de recolección de energía, y *ii*) controlando las funciones del registrador que deben encontrarse operativas en cada instante. Este bloque consta de un regulador DC-DC de alta eficiencia combinado con reguladores lineales de baja caída de tensión, para maximizar la eficiencia energética y posibilitar la operación del registrador aún cuando el voltaje de la batería esté por debajo del voltaje de operación del registrador.

Mientras que ciertas características de la invención se han ilustrado y descrito aquí, muchas modificaciones, sustituciones y cambios podrán ser propuestos por personas con experiencia técnica. Por lo tanto debe entenderse que las reivindicaciones adjuntas están destinadas a cubrir todas las modificaciones y cambios que caigan dentro del verdadero espíritu de la invención.

### **Ejemplo de aplicación**

A continuación se describe la presente invención mediante una forma de realización preferida que detalla de manera amplia, haciendo referencia a las figuras presentadas y dando alternativas constructivas, cada una de las partes que la constituyen. Se brinda toda la información necesaria, conjuntamente con las referencias del estado de la técnica, para que la invención pueda ser reproducida por un técnico medio entendido en la materia.

La unidad de sensado comprende un micrófono para detectar los sonidos producidos, una unidad de medición inercial (IMU) para medir los movimientos que realiza la cabeza del animal, un GPS para determinar la posición del animal, un transceptor para comunicaciones de campo corto (NFC) con sus antenas para determinar los animales que se hallan en las proximidades y un transceptor wi-fi. Los sensores, las antenas y la unidad de procesamiento se ubican en la cabeza del animal y utilizan un bozal para su fijación, como se indica la Figura 2. El micrófono utilizado puede ser cualquier tipo conocido de sensor de sonido que tenga características direccionales (capta sólo los sonidos provenientes de su zona frontal) y se ubica en la cara interior de una banda ajustable (que puede ser elástica) del bozal, que se encuentra ubicada en la parte superior de la cabeza del animal (U2). Esta zona permite registrar de manera más simple y efectiva los sonidos producidos durante la alimentación. La IMU, el GPS y el transceptor NFC se hallan ubicados junto con la unidad de procesamiento (U1), situada en la parte posterior de la cabeza, sobre el cuello (Figura 2). Esta ubicación permite mantener conexiones cortas con el micrófono y las antenas, proteger la unidad de registro y el animal naturaliza rápidamente su presencia. La banda ajustable se halla protegida por una cobertura de material impermeable y goma espuma o similar, la cual protege al micrófono de las condiciones ambientales. La direccionalidad de la zona de captación del micrófono combinada con la aislación pasiva a los ruidos ambientes de los materiales que cubren la banda son suficientes para funcionar sin inconvenientes en las condiciones operativas de campo que se encuentra con mayor frecuencia. Las antenas de transmisión del transceptor NFC se ubican en el bozal en U3 (Figura 2), mientras la antena receptora se encuentra en el dispositivo en U1. De este modo, las antenas de transmisión y recepción se encuentran a una distancia de entre uno y dos metros cuando dos animales se montan.

La unidad de procesamiento está compuesta por cuatro procesadores y un conjunto de circuitos electrónicos integrados especializados que se utilizan para implementar las unidades elementales

de la unidad de procesamiento (Figura 14): *i*) la unidad de adquisición y procesamiento de señales, *ii*) la unidad de análisis de actividades y comportamiento, *iii*) la unidad de almacenamiento y comunicaciones y *iv*) alimentación y administración de energía. La unidad de adquisición y procesamiento de señales se implementa a través de tres circuitos electrónicos dedicados a cada una de las señales registradas: sonido, movimiento y posición. La señal de sonido y movimiento de cabeza son analizadas por el procesador 1, el cual analiza el comportamiento alimentario del animal. Las señales de movimiento son procesadas por el procesador 2, el cual corrige los errores sistemáticos y remueve los ruidos de las señales de movimiento. Los movimientos y posiciones del animal son analizadas por el procesador 3, el cual analiza el comportamiento físico individual y el comportamiento grupal del animal. Finalmente, las tareas relacionadas con el análisis del comportamiento general del animal, el almacenamiento, las comunicaciones y la gestión de energía son realizadas por el procesador 4, el cual está a cargo de determinar el estado del animal (combinando la información de cada aspecto del comportamiento del animal), armar los paquetes de información, transmitirlos y comunicarse con otros dispositivos. Como este procesador está a cargo de una de las tareas que más energía consumen (como la comunicación inalámbrica), también se encarga de gestionar la energía disponible en el dispositivo.

Esta implementación de las tareas que se realizan en el dispositivo permite reducir la frecuencia de reloj de los procesadores a valores muy bajos, gracias a la ejecución paralela y concurrente de las tareas, de manera que es posible utilizar los modos de operación de ultrabajo consumo de los procesadores comerciales. De esta manera, se permite una operación de muy bajo consumo, lo cual habilita a la operación continua del dispositivo a partir del uso de energía solar o cualquier fuente alternativa de energía.

El sonido producido por los animales se registra a través de un micrófono electret direccional conectado a los dos canales de un sistema de adquisición integrado (Figura 2). El sistema de adquisición se encuentra implementado en un circuito electrónico integrado de bajo consumo que incluye en cada uno de sus dos canales: un amplificador de ganancia programable, un circuito de polarización de los micrófonos, un control automático de ganancia con ajuste de nivel de ruido y parámetros programables y un convertor analógico-digital de 24 bit de resolución y frecuencia de muestreo programable hasta 96 kHz. El circuito integrado incluye una interfaz digital serie tipo SPI dedicada para la programación de los parámetros de operación y una interfaz digital serie de audio tipo I2S para lectura de los datos. En la implementación se utilizó una frecuencia de muestreo de 2 kHz, con los circuitos de polarización de micrófonos desconectados y el control de automático de ganancia activado sólo en el canal derecho. La ganancia del amplificador del canal izquierdo fue ajustada a un valor fijo de 12 dB. Los datos resultantes fueron leídos por el procesador a través de una interfaz I2S implementada por medio de una rutina dedicada. La programación de los parámetros del circuito de adquisición se realizó a través una interfaz SPI, implementada a través de hardware dedicado disponible en el procesador. Las señales obtenidas son recuantizadas a 8 bits para utilizar aritmética de punto fijo en los algoritmos.

Los movimientos de la cabeza y cuerpo del animal son registrados a través de una IMU. El sistema de adquisición se encuentra implementado en el mismo circuito integrado de bajo consumo que incluye tres acelerómetros, tres giróscopos, tres magnetómetros con una resolución de 16 bits y un ancho de banda de 100 Hz, una unidad de fusión de datos y una unidad de manejo de energía. El circuito integrado incluye una interfaz digital serie tipo I2C dedicada para la programación de los parámetros de operación y la lectura de los datos. El circuito lee los datos de los sensores y los fusiona para corregir los errores y remover los ruidos presentes, utilizando un algoritmo de estimación de mínimos cuadrados (filtro de Kalman). Los datos crudos provistos por

los sensores y/o parámetros característicos del movimiento calculados por la unidad de fusión como los ángulos de Euler, el vector de gravedad y las aceleraciones lineales, entre otros, pueden ser leídos directamente por el procesador. En la implementación se utilizó un algoritmo de estimación que combina las mediciones de los tres sensores (acelerómetros, giroscopios y magnetómetros) para obtener la orientación absoluta del movimiento. La combinación de la información provista por los tres sensores resulta en una mayor robustez a las perturbaciones y precisión de las estimaciones obtenidas. La frecuencia de muestreo de la IMU se fijó en 10 Hz, con los rangos de los sensores limitados a  $\pm 2$  g/s para los acelerómetros y  $\pm 250$  °/s para los giroscopios, de modo de obtener la máxima sensibilidad. De toda la información disponible se leen: *i)* los ángulos de Euler, *ii)* el vector gravedad y *iii)* las aceleraciones lineales a través de una interfaz I2C dedicada del procesador.

La unidad de análisis de actividades alimentarias se implementa en el procesador 1 a través de las rutinas utilizadas para procesar y analizar los datos obtenidos del micrófono y la IMU. Estas rutinas implementan en el microcontrolador los algoritmos de detección y clasificación de eventos y actividades, además de las rutinas necesarias para extraer, clasificar y cuantificar la información indicativa de las eventos individuales. Estas rutinas procesan las señales registradas a medida que las mediciones son registradas utilizando aritmética de punto fijo. En una primera etapa se calcula la envolvente del canal de audio derecho (procesado por el amplificador con control automático de ganancia) a partir del filtrado del valor absoluto de la señal. El cálculo del valor absoluto de la señal se realiza fijando el signo de todos los datos, mientras que el filtrado se realiza utilizando un filtro de respuesta infinita (IIR) de segundo orden implementado utilizando notación de punto fijo con 8 bits de resolución. La información de la envolvente del bloque de sonido obtenida es submuestreada a 100 Hz. El proceso de submuestreo se implementa por software sobre el registro de datos de la envolvente almacenado en la memoria del microcontrolador. Una vez obtenida la señal submuestreada de envolvente del audio se procede a calcular las características de la señal para su clasificación. Las características utilizadas son: forma, duración, simetría y valor máximo en una ventana. La forma de la envolvente se determina a partir de los signos de las pendientes y la duración del evento se obtiene contando la cantidad de muestras mayores que un valor predeterminado.

Por otro lado, el vector de gravedad y los ángulos de Euler calculados por la IMU son utilizados para complementar la información de audio. El vector de gravedad se utiliza para determinar la distancia de la cabeza al suelo a partir del ángulo de la cabeza con respecto al vector de la aceleración gravitacional, mientras que los ángulos de Euler se utilizan para caracterizar los movimientos de la cabeza. Estas señales son utilizadas para identificar la actividad alimentaria realizada por el animal y las aceleraciones se emplea en el análisis del comportamiento individual (nivel de actividad física).

Finalmente, el algoritmo genera las marcas temporales que indican la presencia de potenciales eventos y segmenta las señales. Las marcas temporales se generan a partir de comparar la envolvente con un umbral variante en el tiempo. Una vez detectado un evento (el valor del umbral es menor que la envolvente), se segmenta la envolvente tomando 60 muestras antes y después del instante en que se detectó el evento. Esta información, combinada con el ángulo de la cabeza (vector de gravedad), sus movimientos (ángulos de Euler) y las características temporales de la envolvente (forma, duración, simetría y valor máximo), es utilizada para detectar y clasificar eventos. El clasificador de eventos es implementado como un perceptrón multicapa que clasifica los eventos acústicos detectados en: Mordida, Masticación, Mordida-Masticación, Silencio y Ruido. Además, calcula los parámetros característicos de cada evento individual: la duración y la energía del evento. La energía de cada evento es calculada a partir de la envolvente del canal izquierdo

de audio (procesado por el amplificador con ganancia fija), el cual preserva la energía de la señal registrada. Sólo se calcula la energía de las masticaciones. Esta información, combinada con la cantidad de mordidas y su duración es utilizada por el modelo de consumo para estimar la cantidad de materia seca ingerida por el animal en el periodo de tiempo considerado. Una vez detectados, clasificados y cuantificados los eventos individuales, otra rutina toma los resultados del clasificador de eventos y los organiza en paquetes. Un paquete se define como el conjunto de eventos en un periodo de 15 minutos. Entonces, una vez armados los paquetes se procede a calcular sus estadísticos básicos (número de eventos, tipo de eventos, duración y energía, entre otros) se analiza y cuantifica la actividad alimentaria desarrollada utilizando un sistema borroso de reglas.

La unidad de análisis de comportamiento individual se implementa en el procesador 3 a través de las rutinas utilizadas para procesar y analizar las aceleraciones medidas por la IMU. Estas rutinas implementan los algoritmos de detección y clasificación de actividades y detección de monta, además de las rutinas necesarias para extraer, clasificar y cuantificar la información indicativa de la actividad física individual. Estas rutinas procesan las señales registradas a medida que las mediciones son registradas utilizando aritmética de punto fijo. La señal de aceleración es procesada en paralelo para i) determinar y cuantificar la actividad física desarrollada por el animal y ii) detectar los eventos de monta. Las tareas necesarias para determinar la actividad física involucran la segmentación de la señal de aceleración en bloques de cinco segundos, a los cuales se le calculan las características más relevantes que permitan caracterizar las actividades físicas realizadas por el animal. En esta implementación los atributos temporales son medidas estadísticas de la señal de aceleración (valor medio, varianza, kurtosis, energía, amplitud, máximo, mínimo) pero pueden calcularse otros. Los atributos frecuenciales son los coeficientes que se obtienen mediante la transformada coseno, que posee ventajas para el cómputo y es equivalente a la transformada discreta de Fourier en este caso. También se pueden obtener atributos frecuenciales mediante el uso de bancos de filtros que brinden información por bandas frecuenciales predefinidas. Toda esta información, combinada con la detección de eventos de monta, es utilizada para detectar y clasificar la actividad física realizada por el animal durante el periodo analizado. El clasificador de eventos es implementado como un perceptrón multicapa que clasifica los actividad detectada en: Parada, Echada, Caminando, Corriendo y Quieta. Además, calcula los parámetros característicos de cada actividad: la duración y la energía del evento. La energía de cada evento es calculada a partir del módulo de la aceleración. Una vez detectadas, clasificadas y cuantificadas las actividades de los segmentos, otra rutina toma los resultados del clasificador de actividades y los organiza en paquetes. Un paquete se define como el conjunto de actividades en un periodo de 15 minutos. Entonces, una vez armados los paquetes se procede a calcular sus estadísticos básicos.

La monta de un animal se detecta a partir de la detección de patrones característicos de aceleración en la dirección contraria del vector de gravedad, de esa manera se evitan errores que pueden producirse por desplazamientos de los ejes de referencia asociados a la IMU. Esta detección se realiza utilizando un filtro de coincidencias. Una vez que se ha detectado la monta (un cambio rápido y de gran amplitud en la aceleración vertical), el procesador 3 se comunica con el otro animal a través de un transceptor NFC que le permite leer el código de identificación del animal montado y transferir su propio código al otro animal. Si el animal montado acepta la monta, el animal que monta no se moverá de modo que la aceleración vertical del animal que monta es muy pequeña (casi nula), con lo cual se confirma la monta y se almacena el código de identificación del animal montado. Este código se transmite al sistema, junto con la información del animal, para su integración con el resto de la información del rebaño.

Los animales que se encuentran en la vecindad del animal se determinan utilizando un transceptor NFC. El sistema de comunicación de NFC se encuentra implementado un circuito integrado especializado que incluye una la gestión automática de las comunicaciones, a nivel capa física y protocolo, de los estándares ISO/IEC 7816, ISO/IEC 14443 y EMVCo 4.3, entre otros. El dispositivo Incluye la interfaz de radiofrecuencia, generación de reloj, el control automático de los voltajes de operación y protección de la interfaz de radio frecuencia. En esta implementación el sistema de comunicaciones utiliza el estándar ISO/IEC 14443 con el protocolo NFCIP-1 en modo activo con una velocidad de transferencia de 424 Kbit/s.

La unidad de análisis de comportamiento grupal también se implementa en el procesador 3, debido a que utiliza información similar a la de comportamiento físico, a través de las rutinas utilizadas para procesar y analizar la información de posición del animal relativa al grupo. Estas rutinas implementan los algoritmos para detectar los animales que se encuentran en la vecindad del animal utilizando un transceptor NFC. Este transceptor permite comunicarse con otro transceptor a una distancia no mayor a los 3 m. Las tareas necesarias para determinar el comportamiento grupal involucran la detección e identificación de los animales que se hallan próximos (a menos de 3 mts) a él y su estado. Los animales en la vecindad son identificados explorando su entorno con el transceptor NFC, el cual establece comunicación con los dispositivos que se hallen próximos a él, y obteniendo el código de identificación de los animales a partir de la comunicación activa entre los transceptores utilizando el protocolo NFCIP-1 del estándar ISO/IEC 14443. El estado es determinado a partir de la búsqueda en una lista de animales en celo provista por el sistema a través del sistema de comunicaciones. Luego, utilizando un sistema basado en reglas borrosas se determina la posibilidad de que el animal se encuentre en celo.

La unidad de análisis de comportamiento reproductivo se implementa en el procesador 4 a través de las rutinas utilizadas para componer y analizar las informaciones generadas por las unidades de análisis de comportamiento alimentario, individual y grupal respectivamente (similar a lo detallado en la Figura 12). Se comunica con los otros procesadores a través de una interfaz SPI de alta velocidad dedicada, la cual permite transferir la información necesaria para la ejecución del algoritmo, mientras mantiene un diseño simple en la placa PCB. Estas rutinas implementan los algoritmos de detección y clasificación del estado reproductivo del animal. Estas rutinas procesan la información provista por los otros procesadores utilizando un sistema experto basado en reglas borrosas. Periódicamente se genera un reporte que es almacenado en el almacenamiento no volátil y transmitido a la estación base a través de un transceptor WiFi de baja potencia. La memoria no volátil del dispositivo del dispositivo está constituida por al menos 16 GB de tipo flash de bajo consumo.

El módulo de comunicaciones puede implementar conexiones punto a punto con una estación base u otros dispositivos para implementar redes con topología estrella o mesh, de acuerdo a las necesidades, con velocidades de transmisión de hasta 250 kbit/s y un alcance de 500 metros a campo abierto implementando el estándar IEEE 802.15.4. Tanto los factores de forma como el consumo de los módulos son pequeños y compatibles con los requerimientos de autonomía del dispositivo. El procesador se comunica con la memoria no volátil a través de una interfaz SPI, implementada con hardware dedicado disponible en el procesador. Por otro lado, el módulo de comunicación se comunica con el procesador a través una interfaz serie asíncrona tipo UART, implementada con hardware dedicado disponible en el procesador.

Finalmente, un bloque de alimentación es el encargado de generar las diferentes líneas de alimentación utilizadas por el dispositivo a partir de una única fuente de energía y administrar la energía disponible en las baterías para maximizar el tiempo de operación del registrador. Esta tarea se realiza de dos maneras complementarias: i) asegurando la máxima carga posible de la



batería a partir de la energía provista por un panel solar y ii) controlando las funciones del registrador operativas en cada instante. Este control se lleva a cabo deshabilitando funcionalidades del registrador de acuerdo a sus prioridades. Es decir, que en la medida en que la carga de la batería disminuya, la frecuencia del uso de las comunicaciones inalámbricas se disminuye, hasta anularse en casos extremos, con el objeto de garantizar el registro y análisis de las actividades del animal. Este bloque es el encargado de gestionar el funcionamiento de los módulos, un cargador de baterías que utiliza energía solar y un regulador DC-DC de alta eficiencia combinado con reguladores lineales de baja caída de tensión, para maximizar la eficiencia energética de este bloque y posibilitar la operación del registrador aún cuando el voltaje de la batería esté por debajo del voltaje de operación del dispositivo.

### Referencias

1. Clapham W, Fedders J, Beeman K and J Neel. "Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle", *Computers and Electronics in Agriculture*, vol. 76, pp. 96-104 (2011).
2. Milone D, Galli J, Cangiano C, Rufiner H and E Laca. "Automatic recognition of ingestive sounds of cattle based on hidden Markov models", *Computers and Electronics in Agriculture*, vol. 87, pp. 51 - 55 (2011).
3. Navon S, Mizrach A, Hetzroni A and E Ungar. "Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring", *Biosystems Engineering*, vol. 114(4), pp. 474-483 (2013).
4. Chelotti J, Vanrell S, Milone D, Utsumi S, Galli J, Rufiner H and L Giovanin. "A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle", *Computers and Electronics in Agriculture*, vol. 127, pp. 64-75 (2016).
5. Nielsen L, Pedersen A, Herskin M and L Munksgaard. "Quantifying walking and standing behaviour of dairy cows using a moving average based on output from an accelerometer", *Applied Animal Behaviour Science*, vol. 127(1), pp. 12-19 (2010).
6. Qasem L, Cardew A, Wilson A, Griffiths I, Halsey L, Shepard E, Gleiss A and R Wilson. "Tri-axial dynamic acceleration as a proxy for animal energy expenditure; should we be summing values or calculating the vector?", *PLoS One*, vol. 7(2), e31187 (2012).
7. Diosdado J, Barker Z, Hodges H, Amory J, Croft D, Bell N and E Codling. "Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system", *Animal Biotelemetry*, vol. 3(1), pp. 3-15 (2015).
8. Arcidiacono C, Porto S, Mancino M and G Cascone. "A threshold-based algorithm for the development of inertial sensor-based systems to perform real-time cow step counting in free-stall barns", *Biosystems Engineering*, vol. 153, pp. 99-109 (2017).

## ***Reclamaciones***

La presente invención reivindica los siguientes aspectos:

1. Un método para detectar celo en animales rumiantes que comprende las siguientes etapas:
  - Detectar, clasificar y cuantificar la actividad física realizada por un rumiante a partir de la medición de los movimientos del animal durante un período de tiempo a través de una unidad inercial de medición ubicada en su cabeza,
  - Detectar, clasificar y cuantificar la actividad alimentaria realizada por dicho animal a través de la medición y análisis de los sonidos o vibraciones producidas por el animal y la posición y movimientos de la cabeza;
  - Detectar, clasificar y cuantificar el comportamiento del animal en el contexto del grupo al que pertenece; e
  - Identificar el comportamiento anormal indicativo de celo en dicho animal al componer los comportamientos previamente identificados y compararlos con el comportamiento típico.
2. Un método de acuerdo con la reclamación 1, que comprende además el establecimiento de una base de datos de comportamiento típico, en el que dicha base de datos se actualiza continuamente.
3. Un método de acuerdo con la reclamación 1, que comprende además la etapa de transmitir al receptor remoto una lista dada el movimiento detectado, la información relacionada con los períodos de la alimentación, la información del comportamiento grupal y la identificación del estado del animal.
4. Un método de acuerdo con la reclamación 1, que comprende además una etapa para enviar la notificación cuando el animal se encuentra en celo.
5. Un método de acuerdo con la reclamación 1, en donde al menos uno de los pasos se lleva a cabo dentro de un dispositivo que se coloca en la cabeza de modo que la inclinación de la cabeza del animal puede ser medida.
6. Un método de acuerdo con la reclamación 1, donde el paso de medir los datos indicativos de la actividad alimentación comprende una etapa de medición de la inclinación con respecto al suelo y movimientos de la cabeza del animal.
7. Un método de acuerdo con la reclamación 1, donde el paso de medir los datos indicativos de la actividad alimentaria comprende el registro, procesamiento y análisis de los sonidos y vibraciones producidas por el animal cuando se alimenta.
8. Un método de acuerdo con la reclamación 1, en el que la etapa de detectar los cambios de comportamiento comprende la medición y procesamiento de las aceleraciones traslacionales, las velocidades de rotación y la medición de los ángulos con respecto al norte magnético de la tierra en hasta tres direcciones ortogonales.
9. Un método de acuerdo con la reclamación 1, en el que la etapa de detectar e identificar la monta de un animal a otro comprende la medición y procesamiento de las aceleraciones traslacionales verticales y la identificación del animal montante a través de un sistema de comunicación de campo próximo.
10. Un método de acuerdo con la reclamación 1, en el que la etapa de detectar el comportamiento grupal de los animales comprende la identificación de las hembras en celo y toros que se hallan a una distancia dada de la hembra considerada a través un sistema de comunicación de campo próximo y dispositivos GPS.
11. Un dispositivo para la detección de celo en una hembra rumiante, que comprende:



- Al menos una unidad de medición inercial (IMU) y GPS para medir el movimiento y posición de dicho animal durante un período de tiempo, en el que el nivel de actividad se indica a través de parámetros indicativos de la señal de movimiento resultante de la composición de las señales producidas por la IMU;
  - Al menos un sensor de sonido o vibraciones para detectar durante un período de tiempo, los datos indicativos de la actividad alimentaria realizada por dicho animal;
  - Un transceptor para comunicaciones de campo próximo (NFC) para identificar los animales que se encuentren a menos de una distancia determinada del animal considerado;
  - Al menos un sistema embebido (basado en microprocesadores, FPGA o ASIC) para procesar las señales adquiridas, componer las señales de la IMU para corregir errores y mejorar la precisión de los datos resultantes de posición y movimiento, extraer nivel de actividad típico de dicho animal en base a dichos datos de movimiento e identificar el comportamiento anormal indicativo de celo a partir del análisis de la parámetros indicativos de los datos de movimiento.
  - Al menos un sistema embebido (basado en microprocesadores, FPGA o ASIC) para procesar las señales sonoras (o vibraciones) adquiridas, filtrar los ruidos y/o perturbaciones, y extraer la información relevante sobre la actividad alimentaria, componer la señales sonoras y la información de la posición de la cabeza para detectar la actividad alimentaria e identificar y cuantificar el comportamiento alimentario a partir del análisis de la parámetros indicativos de los sonidos, la posición y los movimientos de la cabeza.
  - Al menos un sistema embebido (basado en microprocesadores, FPGA o ASIC) para procesar las señales adquiridas a través de la NFC, extraer información de comportamiento grupal típico de dicho animal en base a los datos de posición de los animales e identificar el comportamiento anormal indicativo de celo en el animal al comparar los datos de composición grupal recientemente identificados con los datos de composición grupal típicos extraído.
  - Al menos un sistema embebido (basado en microprocesadores, FPGA o ASIC) para procesar la información generada por los otros bloques, extraer información de comportamiento típico de dicho animal en base a la información generada por los bloques antes mencionados e identificar el comportamiento anormal indicativo de dicho celo en dicho animal.
  - Al menos un sistema de almacenamiento no volátil para almacenar la información generada por los sistemas de procesamiento y análisis de información implementados en el dispositivo.
  - Al menos una unidad transceptora para transmitir el estado del animal, y/o cualquier otra información obtenida a partir de los datos procesados, y recibir información del transmisor remoto, como parámetros de operación y cuándo otros animales han sido identificados en celo.
12. Una fuente de alimentación capaz de obtener energía del sol a partir de uso de paneles solares, o cualquier otra fuente de energía alternativa, de almacenarla y hacer un uso eficiente de la misma a partir del uso de dispositivos de carga y regulación de alto desempeño.
13. Un dispositivo de acuerdo con las reclamaciones 11, que se implementa utilizando una ejecución concurrente de las tareas de modo de reducir al mínimo necesario la frecuencia de operación del dispositivo con la consecuente reducción del consumo de energía, que combinado con una fuente de energía de acuerdo con la reclamación 12 permitan la operación continua del dispositivo para el monitoreo continuo de los animales.

**Figuras**

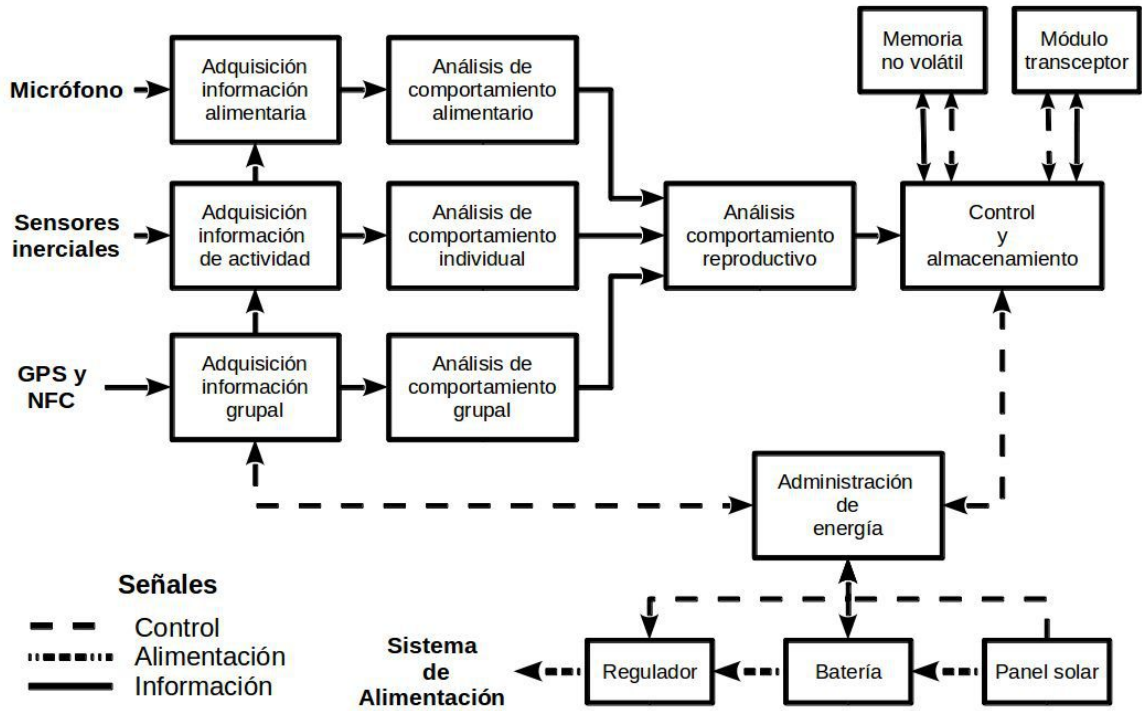


Figura 1

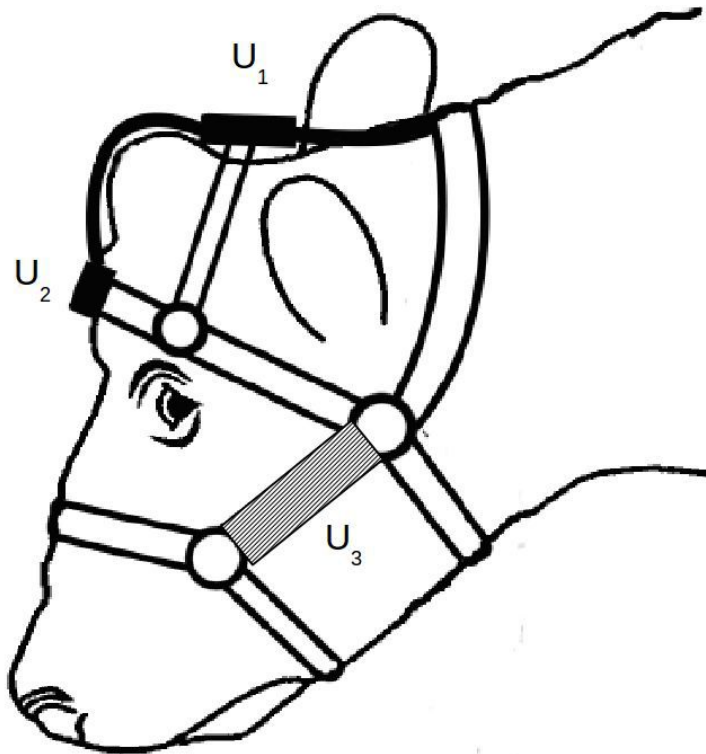


Figura 2

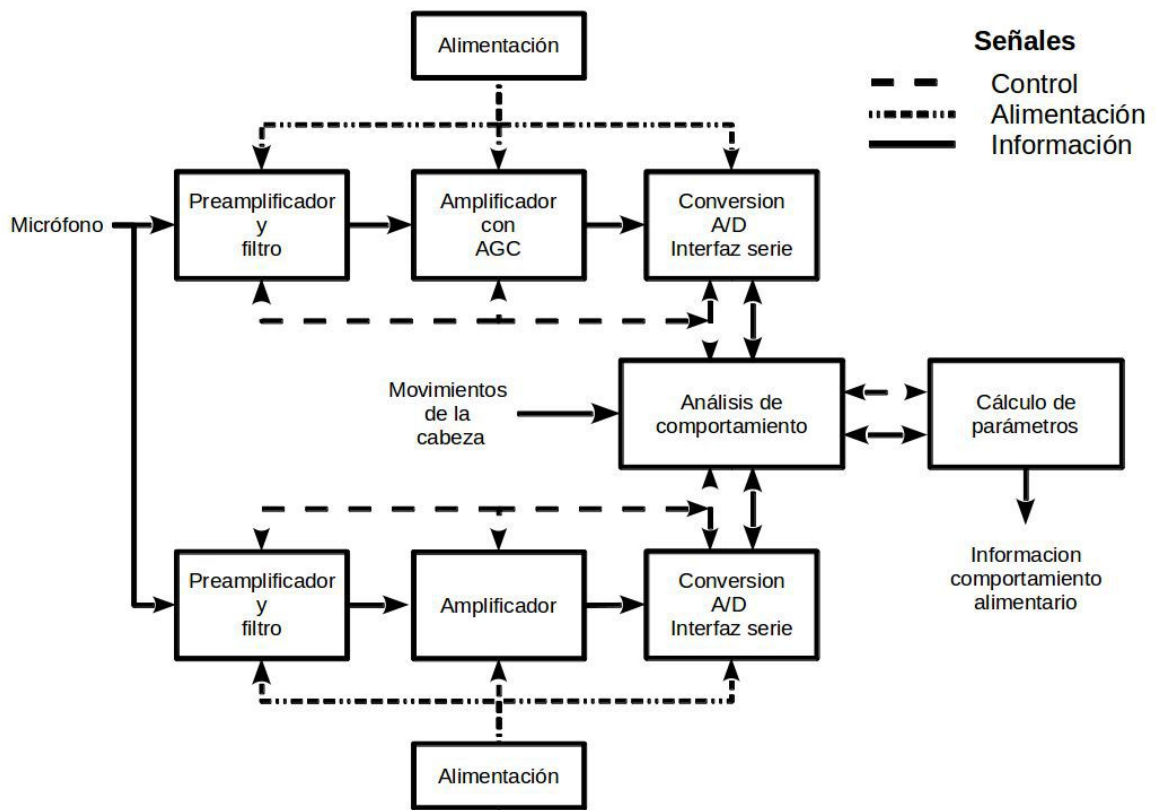


Figura 3

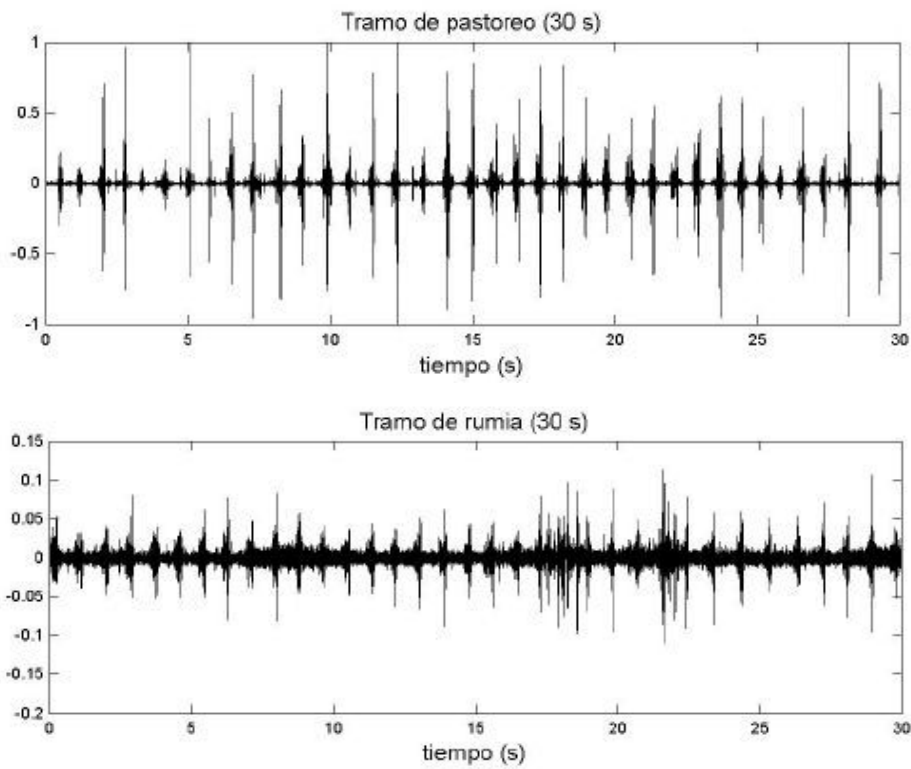


Figura 4

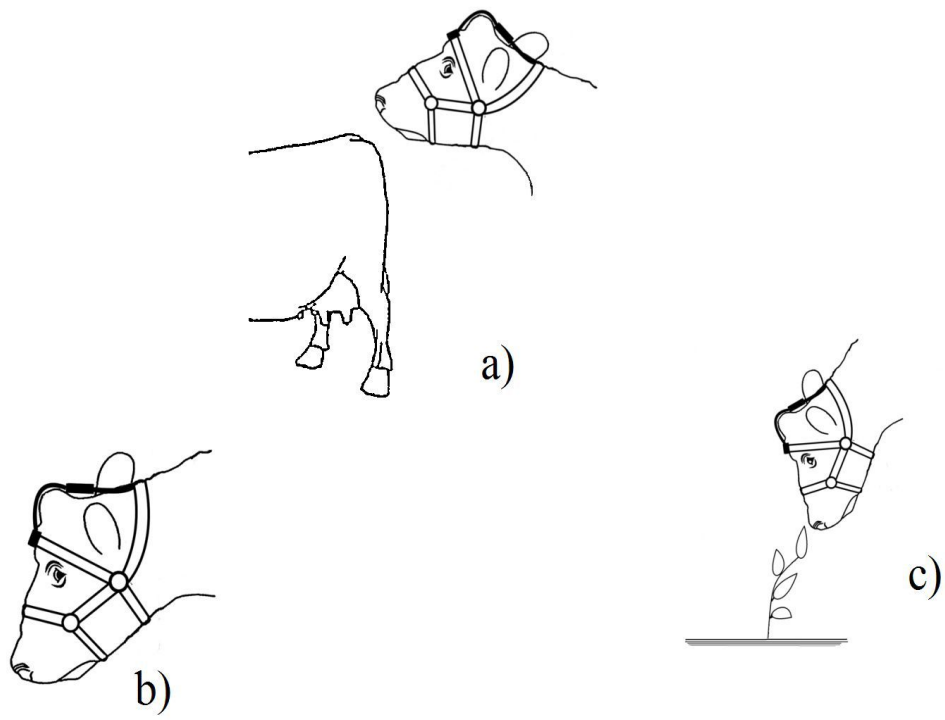


Figura 5

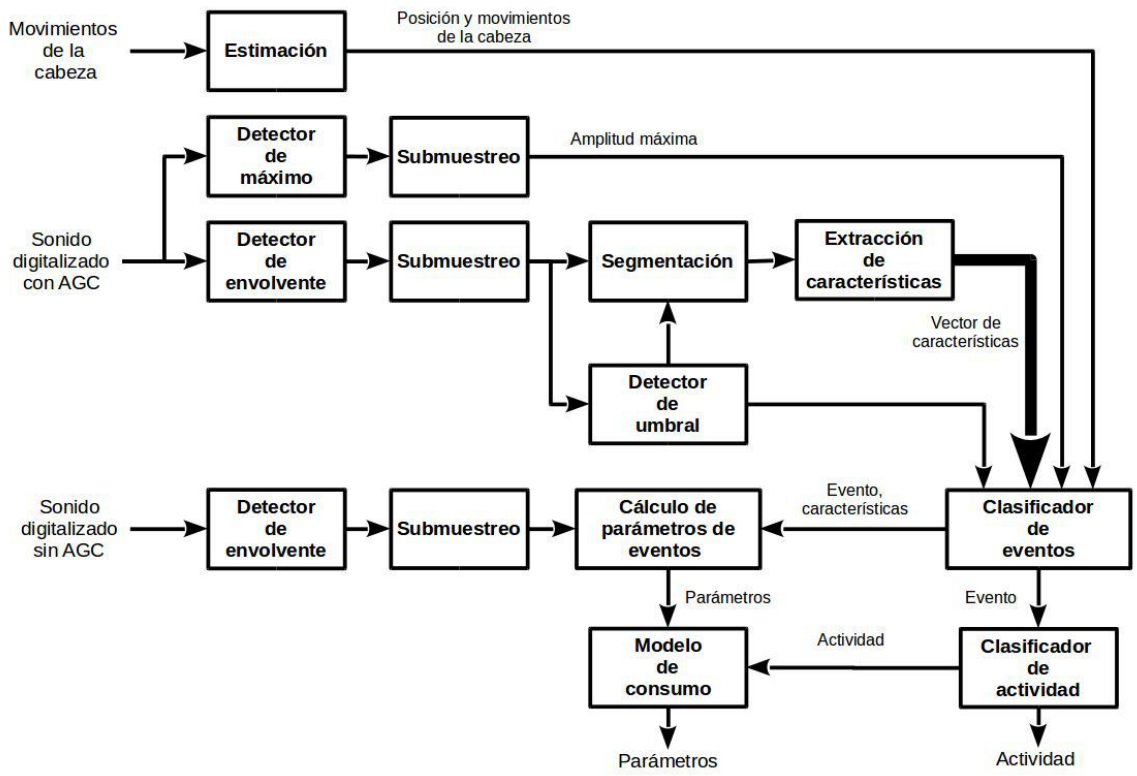


Figura 6

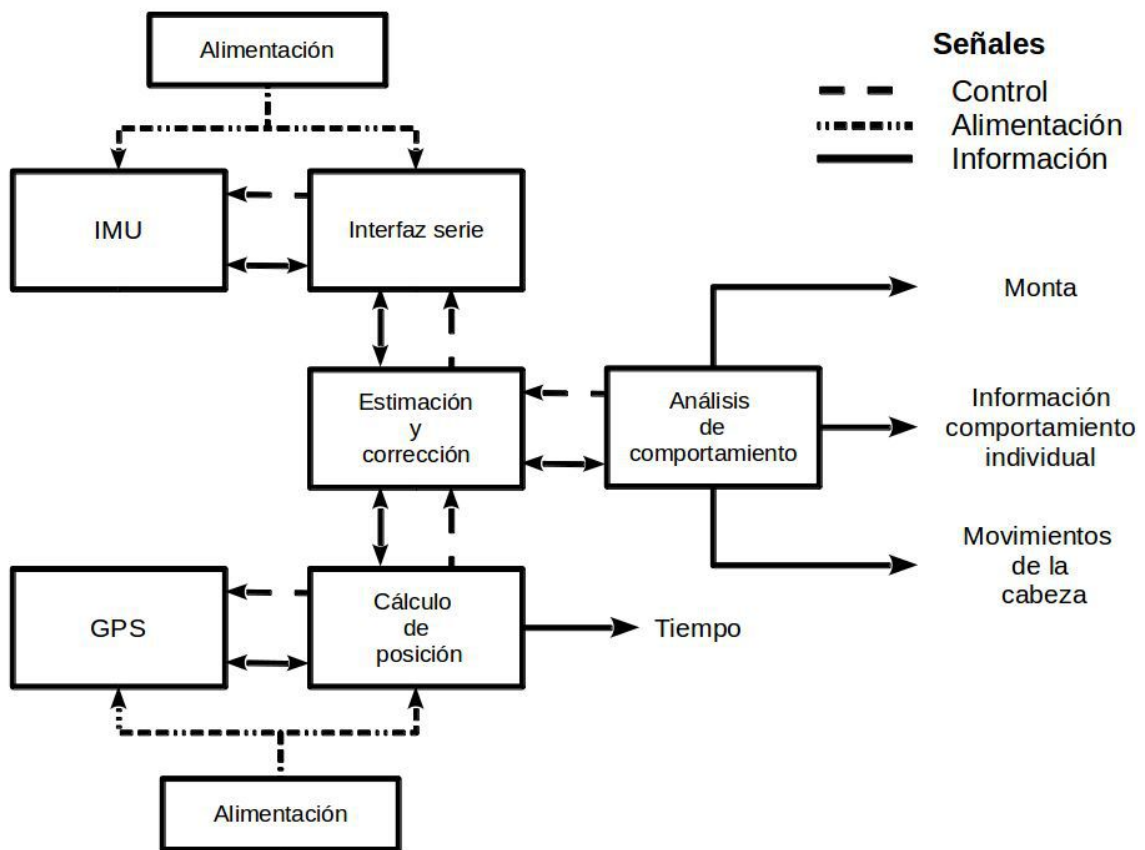
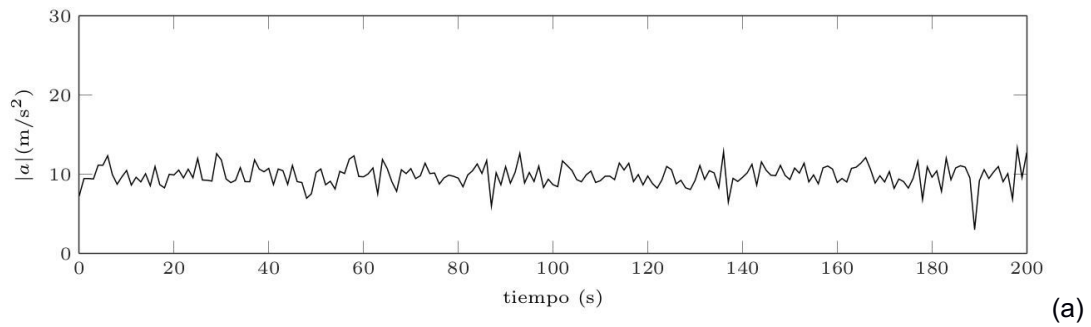


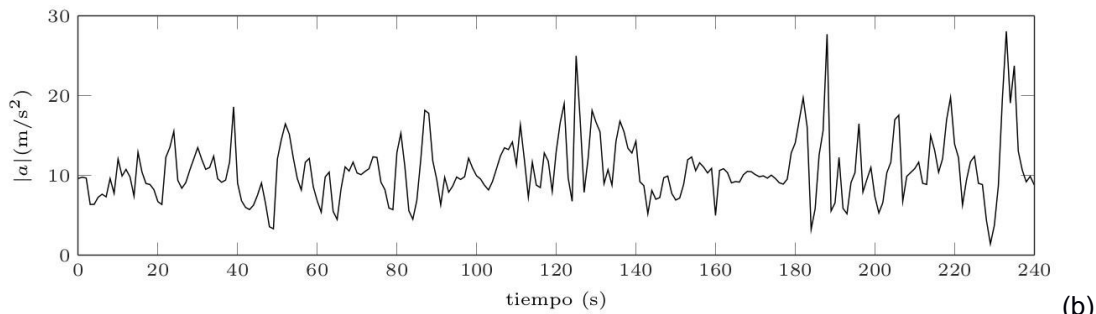
Figura 7

Período de actividad normal



(a)

Período de celo



(b)

Figura 8

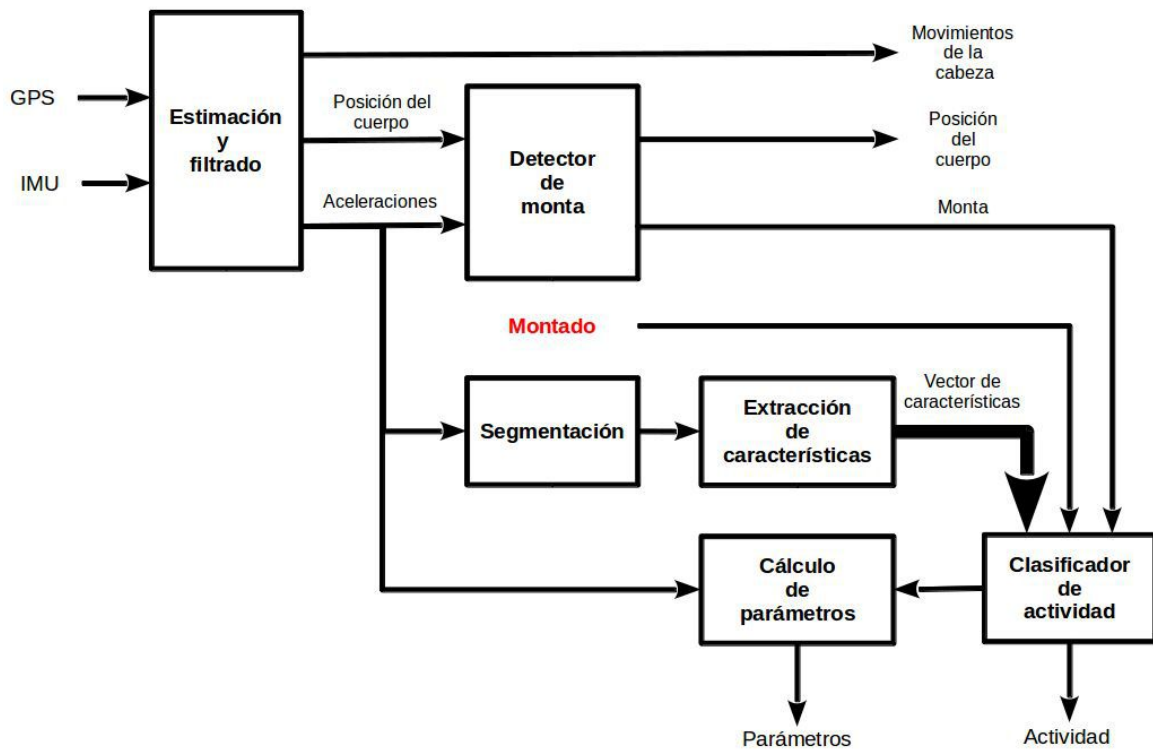


Figura 9

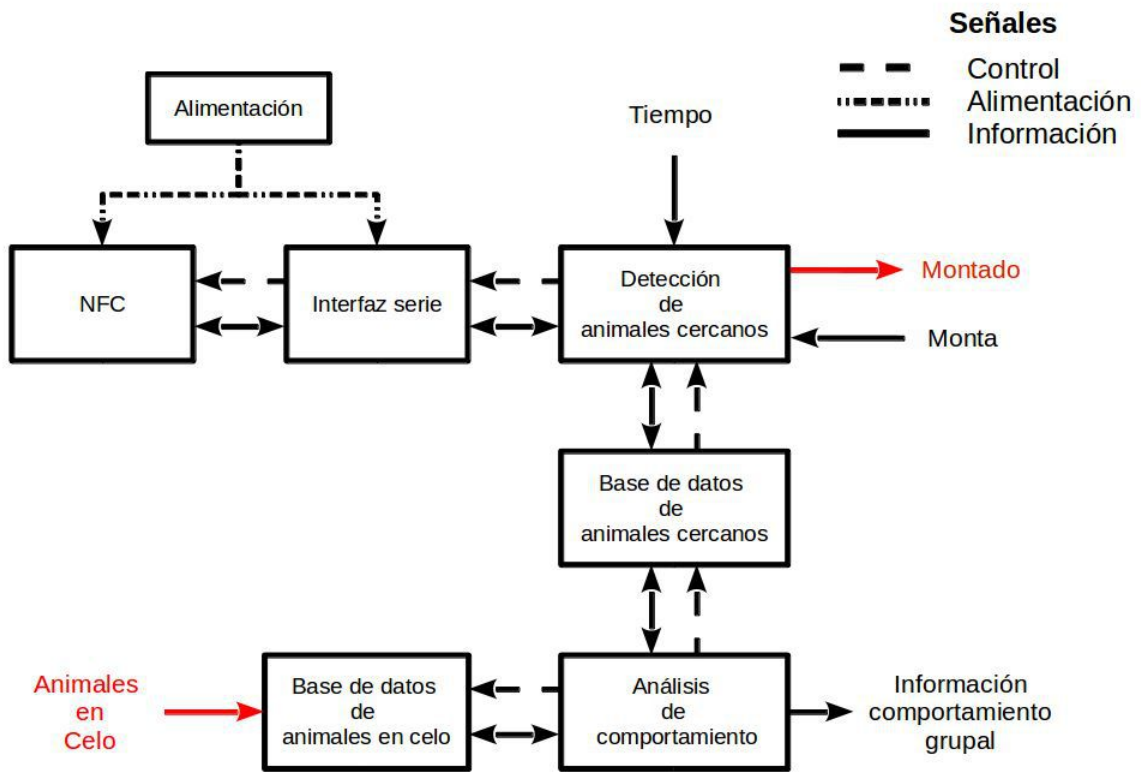


Figura 10

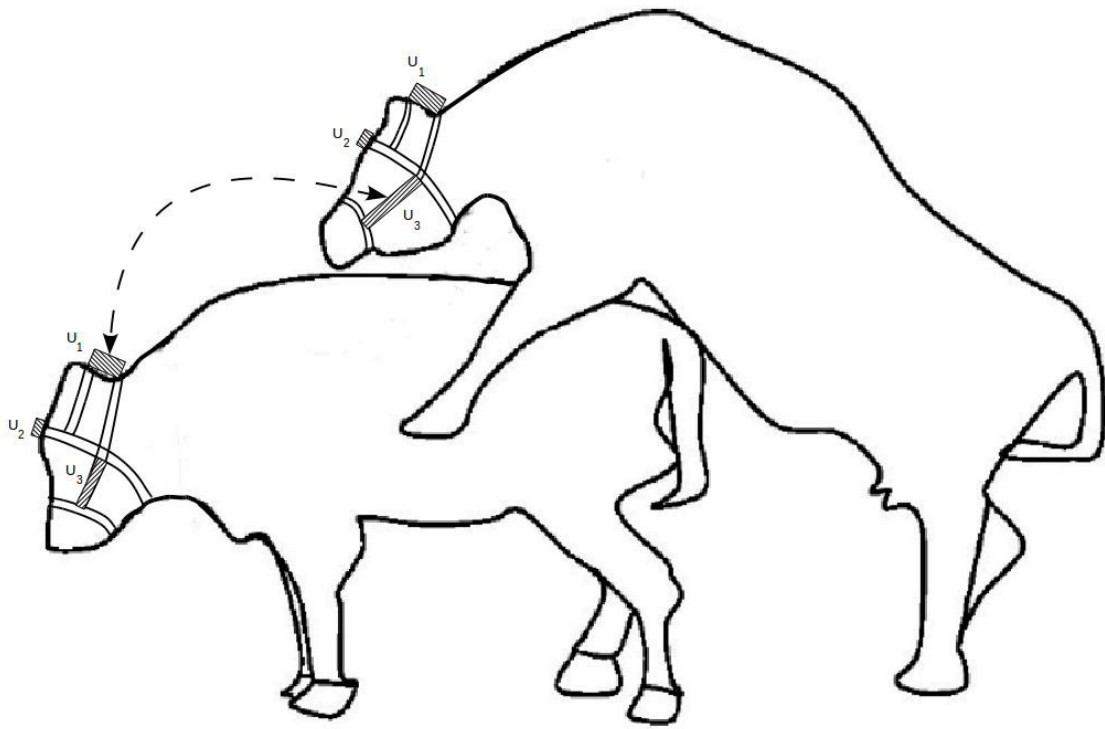


Figura 11

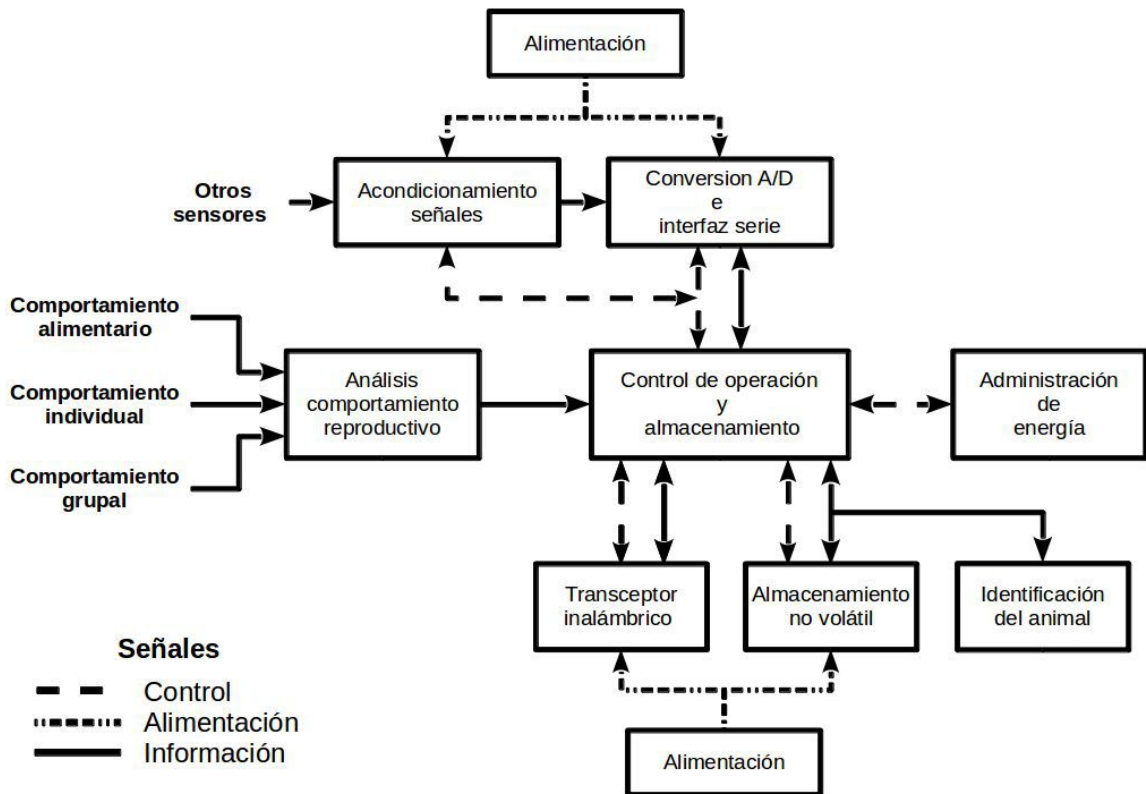


Figura 12



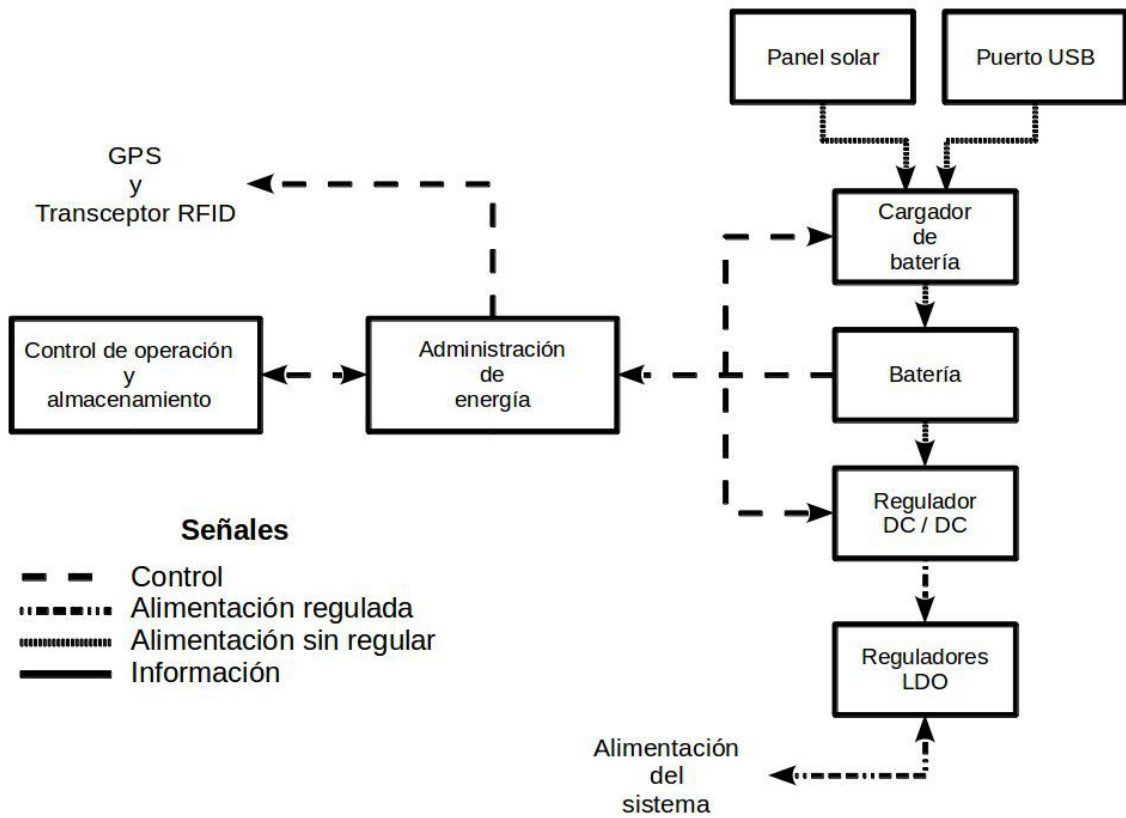


Figura 13

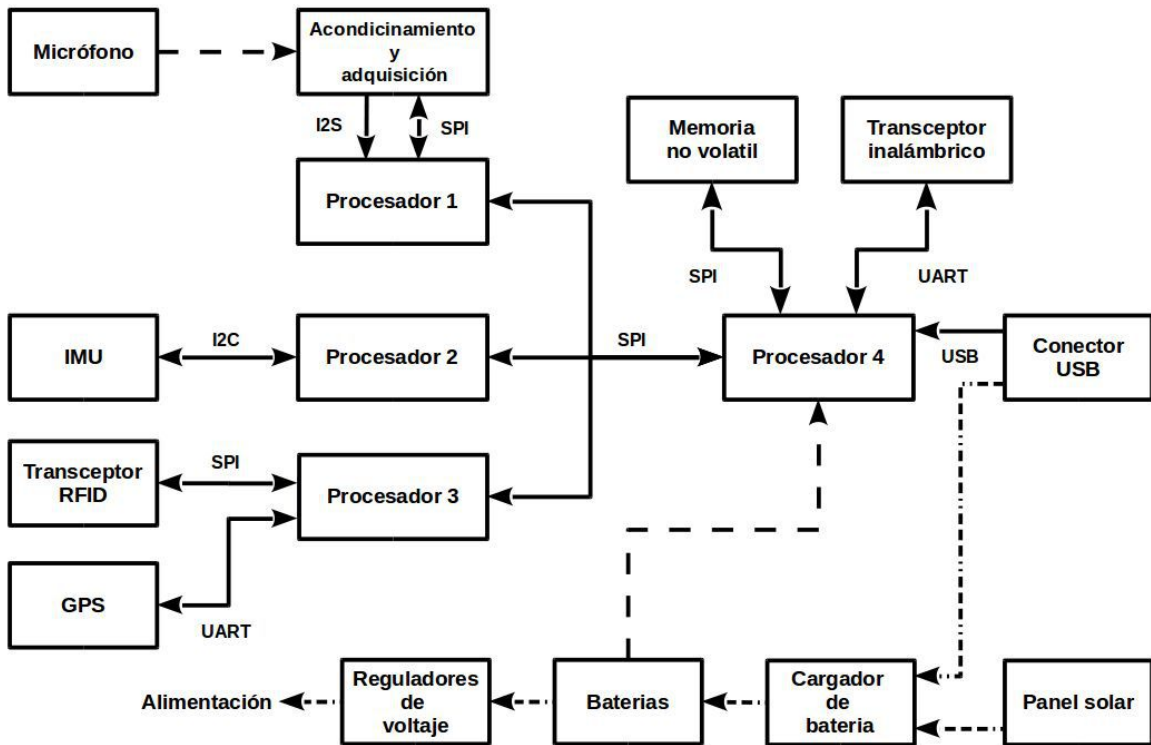


Figura 14



## ***Resumen***

Un procedimiento para el monitoreo y cuantificación del estado reproductivo de rumiantes rumiantes a partir de la detección y cuantificación del comportamiento físico individual, el comportamiento alimentario y el comportamiento grupal del animal y un dispositivo para realizar dicho procedimiento.



# **Doctorado en Ingeniería**

mención en Inteligencia Computacional, Señales y Sistemas

Título de la obra:

## **Reconocimiento de actividades a partir de señales inerciales y acústicas**

Autor: Sebastián Rodrigo Vanrell

Lugar: Santa Fe, Argentina

Palabras Claves:

reconocimiento de actividades,  
procesamiento de señales,  
aprendizaje maquina,  
análisis cepstral,  
ganadería de precisión,  
monitoreo acústico.