A robust computational approach for jaw movement detection and classification in grazing cattle using acoustic signals

Luciano S. Martinez Rau[a], José O. Chelotti[a], Sebastián R. Vanrell[a], Julio R. Galli[b,c], Santiago A. Utsumi[d], Alejandra M. Planisich[c], H. Leonardo Rufiner[a,e], Leonardo L. Giovanini[a]

[a] *Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL/CONICET, Argentina*
[b] *Instituto de Investigaciones en Ciencias Agrarias de Rosario, IICAR, UNR-CONICET, Argentina*
[c] *Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Argentina*
[d] *W.K. Kellogg Biological Station and Departament of Animal Science, Michigan State University, United States*
[e] *Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina*

Abstract

Monitoring behaviour of the grazing livestock is a difficult task because of its demanding requirements (continuous operation, large amount of information, computational efficiency, device portability, precision and accuracy) under harsh environmental conditions. Detection and classification of jaw movements (JM) events are essential for estimating information related with foraging behaviour. Acoustic monitoring is the best way to classify and quantify ruminant events related with its foraging behaviour. Although existing acoustic methods are computationally efficient, a common failure for broad applications is the deal with interference associated with environmental noises. In this work, the acoustic method, called Chew-Bite Energy Based Algorithm (CBEBA), is proposed to automatically detect and classify masticatory events of grazing cattle. The system incorporates computations of instantaneous power signal for JM-events classification associated with chews, bites and composite chew-bites, and additionally between two classes of chew events: i) low energy chews that are associated with rumination and ii) high

Email address: `lmrau@sinc.unl.edu.ar` (Luciano S. Martinez Rau)

energy chews that are associated with grazing. The results demonstrate that CBEBA achieve a recognition rate of 91.9% and 91.6% in noiseless and noisy conditions, respectively, with a high classification precision and a marginal increment of computational cost compared to previous algorithms, suggesting feasibility for implementation in low-cost embedded systems.

Keywords: Acoustic monitoring, Cattle grazing behaviour, Jaw movement classification, Noise robustness, Pattern recognition, Sound energy analysis.

1. Introduction

Precision livestock farming typically integrates smart animal monitoring technologies to aid farmers with relevant management decisions regarding animal nutrition, health and welfare (Michie et al., 2020). The deployment of animal monitoring dashboards has been enhanced recently by improved sensors (Andriamandroso et al., 2016), advanced communication technologies and enhanced visualization tools allowing for rapid inspection of production traits and behaviours associated with specific activities, changes of location and body posture (Berckmans, 2014).

The use of modern technologies based on fixed video cameras allow for individual or group behaviour monitoring in an automatic, continuous and non-intrusive way in a given fixed area (Fuentes et al., 2020). Their use is limited to small farming areas such as pens and stables. On the other hand, the use of small wearable video cameras on animals would allow to expand the operating region, although their application still needs further development (Saitoh and Kato, 2021). Thus, wearable sensors are the most widely used acquisition method to cover large farm and field areas. However, their operational requirements, primarily device portability, robustness and power capabilities, along with the computational cost and complexity of analytical components often represents an obstacle for further technological progress and adoption (Stone, 2020).

Ones of the most frequently used monitoring techniques are the position and motion sensors, which allows the surveillance of cattle and sheep movements (Andriamandroso et al., 2016). Nose band sensors (Nydegger et al., 2010; Werner et al., 2018; Zehner et al., 2017), multidimensional accelerometers (Andriamandroso et al., 2017; Greenwood et al., 2017; Smith et al., 2016) and jaw recorders have been applied to monitor animal locomotion as well as feeding and rumination activities, being used to alert farmers on behavioural changes associated with diseases, estrus and labor. On the other hand, acoustics methods have been used for monitoring livestock feeding behaviours. Laca et al. (1992) used directional microphones, attached to the forehead of animals, for analysing masticatory sounds in cattle (Galli et al., 2018) and sheep (Galli et al., 2011), as well as for accurate discrimination between feeding and rumination bouts (Chelotti et al., 2020; Vanrell et al., 2018) and for feed intake prediction based on sound energy fluxes (Galli et al., 2018; Laca et al., 2000).

Masticatory sounds are the result of JM-events associated with bites, chews and composite chew-bites (Laca et al., 1992). A grazing bite includes the apprehension and severance of herbage. Chews include the crushing, grinding and processing of herbage during ingestion or rumination. Finally, chew-bites include the combination of chewing and biting in the same JM. Thus, the attributes and statistics of JM-events provide a reliable measure for identification of grazing (include bites, chews and chew-bites) and rumination (chews only) activities and related behaviour events (Chelotti et al., 2016; Milone et al., 2012). Rumination frequency is related to digestion processes and serves as indicator of a suitable rumen health (Sauvant, 2000). The sound properties of bites and chew-bites typically relate to common plant traits and feed structural characteristics (Laca et al., 2000), providing insights of short-term intake rate (Galli et al., 2018) and daily grazing time (Chelotti et al., 2020), two major determinants of the daily feed intake (Hodgson, 1990). A declining rumination time has been shown to be correlated with a declining feed intake (Watt et al., 2015), an acute stressors (Schirmann et al., 2011), an onset of diseases (DeVries et al., 2009), as well as the beginning of estrus (Schirmann et al., 2009) and parturition (Schirmann et al., 2013).

The acquisition of masticatory sounds is the first step of a good acoustic method because registered signals require further processing, analysis and information weighing, and extraction to become useful and insightful for animal monitoring. The analysis of masticatory sounds has been significantly improved in recent years. Milone et al. (2012, 2009) used concepts from automatic speech recognition to develop an algorithm based on hidden Markov models capable of identifying chew, bite, and chew-bite in sheep and cattle. It combines spectral analysis with language-based analysis to detect and classify JM-events. These algorithms achieved an average recognition rate of 80% of successful classifications of bites, chews and chew-bites when tested in controlled experiments lasting a few minutes under good signal-to-noise ratio (SNR) conditions. Chelotti et al. (2016) proposed an alternative algorithm based on time-domain features of sound signals, achieving similar JM-events recognition rate success at a lower computational cost, compared to Milone et al. (2012), such that it can be implemented in portable microcontroller-based embedded systems (Deniz et al., 2017). More recently, Chelotti et al. (2018) modified and improved this algorithm (Chelotti et al., 2016) using concepts and tools derived from signal processing, pattern recognition and artificial intelligence areas without significantly increasing the computational cost. This algorithm, called Chew Bite Intelligent Algorithm (CBIA), attenuates the effects of time-varying noises and trends, and it achieves a 90% recognition rate. Although CBIA showed good performance for moderate SNR, significant limitations arose when it was employed in farming and husbandry environments, which typically involved louder and time-varying noise and disturbance sources. These negatively affect the detection, features extraction and classification of JM-events since they can not be completely removed from processes internal signals (see Chelotti et al., 2018).

The present work documented and tested the integration of a new set of tools for pattern recognition analysis and artificial intelligence for robust online analyses of masticatory sound signals collected from grazing livestock. The main objective is to achieve a more robust detection and classification of JM-events than the CBIA algorithm. Specifically, the new algorithm was especially designed to: i) attenuate distorting effects of environmental noises on masticatory signals likely associated with typical farming conditions and animal handling procedures; and ii) improve the detection and classification of JM-events without significantly increasing the computational cost. The computational cost and a cost-benefit analysis of the new algorithm and CBIA were also evaluated to assess future feasibility for real-time execution in low-cost embedded systems.

The paper is organised as follows: Section 2 analyses the CBIA and presents the new classification features intended to attenuate effects of noises on detection and classification tasks. Then, the proposed algorithm is introduced. Also it is described the acquisition of datasets, the performance measures and the experimental setup used to validate the algorithms. Section 3 shows the comparative results for the proposed algorithm and CBIA. A focused discussion and main conclusions follow in Section 4 and Section 5, respectively.

2. Materials and Methods

In the following section, the operation and weakness of the former CBIA are briefly described. Then, the approach of the proposed algorithm is presented. It is based on the use of instantaneous power signal and it incorporates a novel set of features that are more robust to withstand against likely distortions and interferences produced by environmental noises. The algorithm recognises and discriminates JM-events into four classes (Fig. 1): exclusive rumination-chews (RC - chews related to rumination only), exclusive grazing-chews (GC - chews related to ingestive grazing only), exclusive grazing bites (B - bites taken during grazing) and composite chew-bite (CB - compound chew-bite taken during grazing).

2.1. CBIA review

The CBIA is a real-time pattern recognition system that detects and classifies masticatory sounds produced by ruminants into three JM-event classes (chew (C), bite (B) and chew-bite (CB)) using heuristic features (Chelotti et al., 2018). It is characterised by a sequence of generic stages that allows the processing, description and analysis of the sound signal: conditioning, pre-processing, segmentation, feature extraction and classification. In the **signal conditioning stage** the sound signal is conditioned and filtered to improve its SNR using an adaptive low-pass filter. Then, the pre-processing stage computes the sound envelope, from the filtered signal, and decimates it to reduce the computational cost of following stages. The **segmentation stage** identifies the candidate JM-events (i.e. peaks in the envelope signal) using a time-varying threshold. Once a candidate JM-event is detected, the following **features** are extracted from the filtered sound intensity or from the sound envelope: amplitude, duration, shape index and symmetry of the candidate JM-event. Then, the extracted features feed a classifier (**classification stage**) that assigns each candidate JM-event to one of the three possible JM-event classes. Two stages (segmentation and feature extraction) have been specifically designed for CBIA while the remaining ones (conditioning, pre-processing and classification) can be implemented using any algorithm available in the signal processing and computational intelligence literature (Chelotti et al., 2018). This fact allows to implement distinct combinations of algorithms that bring recognition systems with different computational and performance characteristics. The most adequate combination for real-time operation is the algorithm that uses an adaptive least mean square filter (LMS) for signal pre-processing and a multilayer perceptron (MLP) for classification, leading to CBIA LMS-MLP variant.

The CBIA showed recognition rates of ∼90% under controlled experimental conditions (low noise environment). On the same conditions, other known methods achieved lower rates: CBRTA (∼77%) and CBHMM (∼80%). Therefore the CBIA will be used as baseline in this work.

A comprehensive study of the effect of environmental noises and disturbances in CBIA has not been carried out. Incoming sound signals usually have noises and slow varying patterns superimposed on the masticatory sounds produced by animals, causing JM-events detection and classification mistakes. The tasks performed by signal conditioning and pre-processing stages leave some noises and disturbances unaffected. They can distort the sound envelope, upsetting the segmentation and feature extraction stages and hampering the overall algorithm performance. Therefore, to improve robustness in a noisy environment, alternative processing strategies and new features should be used.

2.2. Proposed features

The key problem of the former CBIA is the sensitivity to changes in the sound envelope. Any noise recorded together with masticatory sounds distorts the envelope, leading to errors in the detection stage and misinterpretations of the features decreasing the successful JM-events classification rate.

Since noises overlap with the target signal, they cannot be completely removed from the incoming signals by the use of real-time low-computational-cost signal processing techniques. The alternative tested in this work is to combine pre-processing techniques with the use of the envelope of the instantaneous power signal, instead of using the sound envelope to detect JM-events and extract their corresponding features. The instantaneous power is computed as the scalar product of the conditioned incoming signal at a given point of time. Thus, the envelope of the instantaneous power signal works as an expander, increasing the event detection sensibility. It is analogous to an asynchronous demodulation technique, which is frequently used to improve the SNR for small signals in measurement systems (Roden, 1996).

In this work, we propose to extend computation to the following set of temporal features, extracted either from the instantaneous power signal or its envelope:

- **Sign of the envelope slope**: This feature represents the shape of the JM-event and is calculated as the number of times that the envelope slope changes from positive to negative. Typically, the envelope changes from positive to negative once for GC and B, two or more times for CB, and one or more times for RC (Fig. 1, row 3).

- **Accumulated envelope speed**: The speed of the envelope slope (or absolute magnitude of change) is related to the sound intensity and variations. This feature is computed as the cumulative sum of the speed of the envelope slope. It distinguishes low intensity RC from any other grazing event. Also, a GC is differentiated from a B and a CB, which present similar values between them. The Fig. 1, row 4, shows how the accumulated envelope speed increases as a function of time.

- **Duration**: event duration is the time that an JM-event takes place. A typical compound grazing CB has a greater duration than others grazing events. In turn, the duration of a RC is similar to or longer than a GC due to the moisture content of the chewed material (Fig. 1, row 5) (Galli et al., 2020).

- **Symmetry**: is related to the shape of the event and is a measure of the length symmetry of a JM-event computed at peak envelope signal. The value is greater for B than CB, while GC and RC typically present variable values between B and CB (Fig. 1, row 5).

- **Total energy**: is the accumulated value of the instantaneous power signal and it is related to the intensity and duration of the JM-event. The value is higher for CB than B, and B usually have higher value than GC and RC. In turn, RC have a much lower value than any grazing event because the ruminated herbage has already been partially chewed and crushed. The Fig. 1, row 6, shows how the event energy increases as a function of time.

The sign of the envelope slope and duration features have been already used by Chelotti et al. (2018) but in the present work the same features are extracted from a different signal: the instantaneous power signal.

2.3. The CBEBA description

In this section the proposed algorithm called *Chew-Bite Energy Based Algorithm* (CBEBA) is introduced. CBEBA performs three tasks: i) JM-event

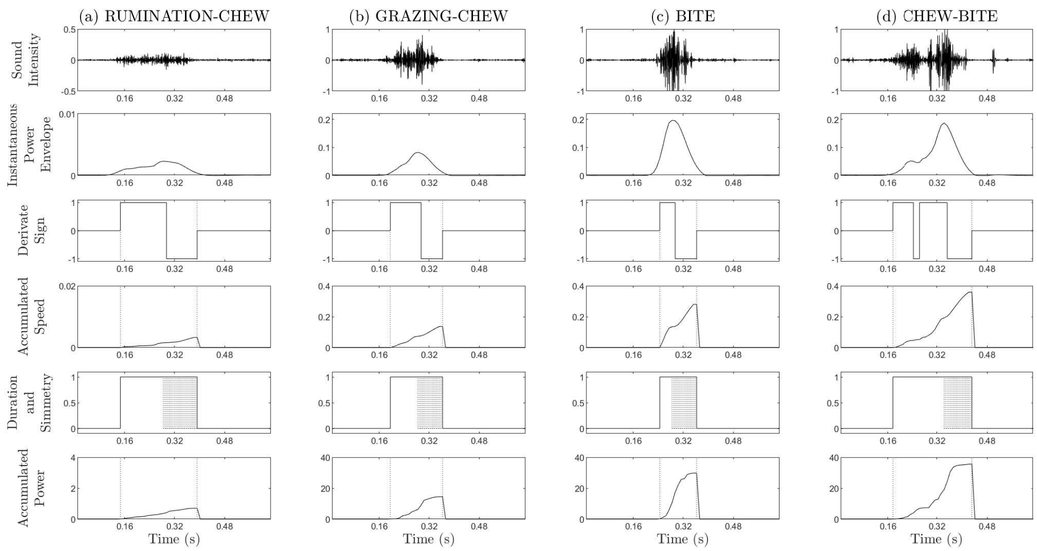

Figure 1: Signals of typical acoustic events produced by jaw movements and their corresponding features for: a) rumination-chew, b) grazing-chew, c) bite and d) chew-bite obtained from the noiseless dataset described in section 2.4. In each row, 1. acoustic signal, 2. instantaneous power envelope, 3. sign of the envelope slope, 4. accumulated envelope slope change rate or speed, 5. duration and symmetry, and 6. accumulated power.

detection, ii) JM-event classification and iii) parameters tuning. JM-event detection is performed by searching for peaks in the instantaneous power signal envelope. Once a possible JM-event has been detected, the features are extracted from the instantaneous power signal and an evaluation is carried out indicating whether the considered event corresponds or not to a true JM. The JM-event classification is carried out using a machine learning technique of low computational cost, to allow future implementation in low-power embedded devices. Finally, parameters tuning is performed. In the time between two consecutive JM-events, the background noise level is estimated and used for the adjustment of internal variables, which significantly improves the algorithm robustness against time-varying conditions.

The online implementation of the algorithm for operation real-time can be internally divided into six stages in a feedback configuration (Fig. 2):

Stage 1 – Signal pre-processing: The input signal usually has noises and disturbances that distort the masticatory sounds produced by animals. Therefore, it is necessary to process the signal in order to remove them. This task is performed by limiting the signal bandwidth with a second-order Butterworth band-pass filter where the energy spectral density of the JM-events is located. The filtered signal is multiplied by itself to obtain the instantaneous power signal ($p(k)$), where k corresponds to the sampling frequency of the input audio signal. The gain of the acquisition system serves as input to normalize the incoming audio signal for further processing, matching its range with the range of the device where the algorithm is running.

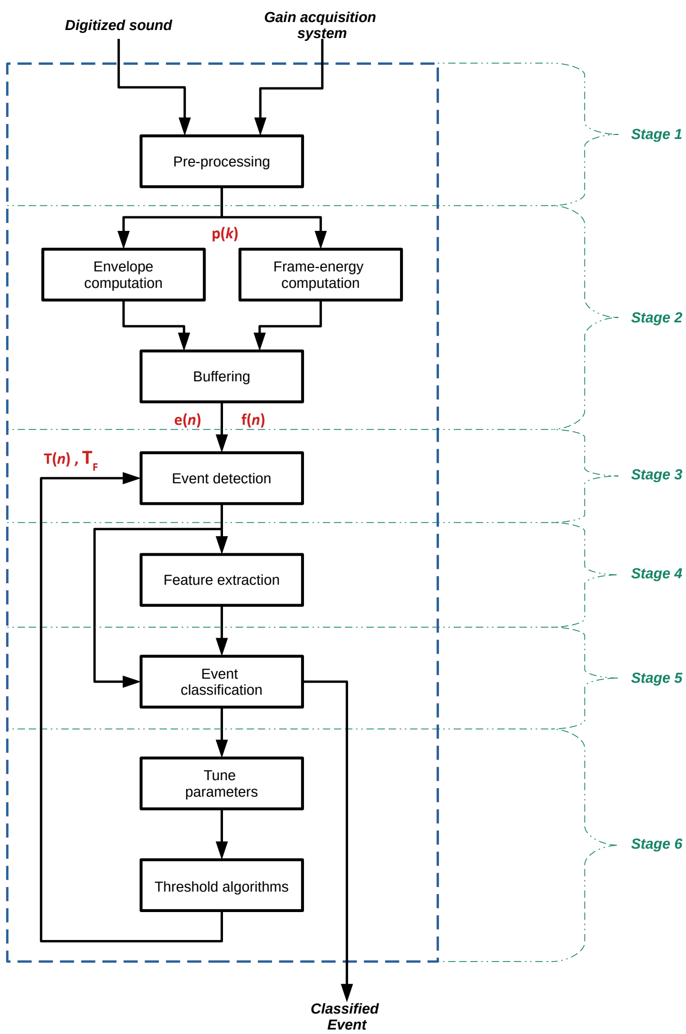

Figure 2: General top-down block diagram of the Chew-Bite Energy Based Algorithm (CBEBA) showing the structure and flow of functions along six operating stages.

Stage 2 – Buffering: The pre-processed instantaneous power signal $p(k)$ follows two parallel paths. In the first path, the envelope computation requires three steps. In the first step, the instantaneous power signal $p(k)$ is filtered with a second-order low-pass Butterworth filter, obtaining the envelope of $p(k)$. In the second step, the low frequency envelope signal is subsampled to reduce the computation cost of the algorithm. Finally, the decimated envelope is saved in the envelope buffer $e(n)$, where n corresponds to the updating frequency of the buffer. In the second path, the frame-energy buffer $f(n)$ stores the energy of $p(k)$ computed by non-overlapping frames, where a "frame" is the cumulative summation of $p(k)$ during a period of time.
Stage 3 – JM-event detection: The buffers $e(n)$ and $f(n)$ contain one second of signals information to be analysed. A peak in $e(n)$ denotes the presence of a possible JM-event. Each peak is detected as a change in the sign of the derivative of the envelope located at the centre position in the e(n). In this case, a second peak is searched beyond the centre position. Each peak must be higher than the value of a given time-varying threshold $T(n)$. If a second peak was found, the middle position of the peaks is centred in $e(n)$.
The time-varying threshold algorithm $T(n)$ is a piecewise linear function that considers both the same anatomical and behavioural characteristics of the animal proposed in the threshold algorithm used in the CBIA and time-varying feeding activities conditions. A description of the parameters involved in the computation of $T(n)$ are available in the supplementary material.
Once a peak has been detected in $e(n)$, the start and end of the candidate JM-event are defined from $f(n)$. The start (n_{START}) and end (n_{END}) of the candidate JM-event correspond to the minimum and maximum n where $f(n) > T_F$ during a hangover period. T_F is the adaptive energy threshold level described in stage 6.
Stage 4 – Features extraction: n_{START} and n_{END} positions of the candidate JM-event are used as boundaries to compute and extract the following features: sign of the envelope slope, accumulated envelope speed, duration, symmetry and total energy, as it was described previously in section 2.2.
Stage 5 – JM-event classification: Once the features have been extracted, the candidate JM-event is analysed to determine if the event should be further classified or not. To be considered a JM, the duration feature must be in a predefined range and the total energy feature must be greater than a certain value of the energy in $f(n)$. Otherwise, the tune threshold algorithm is informed, and the extracted features are discarded. The final classification

step is performed by a MLP.

Stage 6 – Threshold algorithms and parameters tuning: Once the JM-event has been classified, some internal values are extracted from the buffers $e(n)$ and $f(n)$. These values are used to update parameters associated with the time-varying threshold algorithm $T(n)$ and the adaptive energy threshold algorithm T_F. The computation of internal values and parameters related to both algorithms are available in the supplementary material.

2.4. Datasets

Two independent datasets of dairy cow sound records were used to implement and evaluate the proposed algorithms. The first dataset (referred as DS1) was acquired during grazing feeding trials performed at the dairy farm of the W.K. Kellogg Biological Station of Michigan State University, United States, 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, as described in Vanrell et al. (2018). Cows were housed and managed on a pasture-based robotic milking system with voluntary cow traffic as described previously in Watt et al. (2015). During six non-consecutive days the foraging behaviour of five lactating multiparous Holstein cows weighing 652 ± 40 kg was continuously monitored. The dairy cows were group grazed on perennial ryegrass (*Lolium perenne*) / white clover (*Trifolium repens*) and orchardgrass (*Dactylis glomerata*) / white clover pastures as part of a larger herd of ∼140 cows.

The second dataset (referred as DS2) was acquired during grazing feeding trials conducted at the dairy farm of the Campo Experimental J. Villarino of Universidad Nacional de Rosario, Argentina, in October 2014. Project protocols were previously evaluated and approved by the Committee on Ethical Use of Animals for Research of the Universidad Nacional de Rosario. The foraging behaviour of five lactating multiparous Holstein cows weighing 570 ± 40 kg grazing on alfalfa (*Medicago sativa*), fescue (*Festuca arundinacea*) and prairie grass (*Bromus catharticus*) mixed pastures, were continuously monitored during six non-consecutive days. The experimental cows were managed along with a larger dairy herd (150 cows) and milked twice a day (∼6 am - 6 pm).

The two field studies used the same cow acoustic halters to record masticatory sounds. Each halter included a directional microphone pressed to the forehead of each animal connected to a digital recorder (Sony Digital ICD-PX312, Sony, San Diego, CA, USA). Microphones were held and protected by an elastic band attached to halters, avoiding microphone movement, friction and scratches. On a given recording day, the five acoustic halters were randomly assigned to each of five cows and were rotated across cows and sampling days. This design allowed to control for any likely device effect (i.e. all cows were tested with all halters). All recordings were saved in waveform audio (WAV) file format, using 16-bit resolution and a sampling rate of 44.1 kHz. The DS1 dataset is composed of eighteen short-term records of ingestive grazing sounds (*grazing-segments*) of 150 s for each one, and eighteen short-term segments of rumination sounds (*rumination-segments*) of 75 s for each one. The selected recordings were captured with minimum or no influence of external environmental noise; thus, represented ideal acoustic conditions for algorithm testing. A typical *grazing-segment* contains more than 170 JM-events whereas a typical *rumination-segment* contains more than 65 JM-events, which results in over 4,200 JM-events in DS1 (32% RC, 18% GC, 9% B and 41% CB). The DS2 dataset contains more than 5,200 JM-events (28% RC, 28% GC, 11% B and 33% CB) collected from sixteen segments of grazing activity and sixteen segments of rumination activity of same duration as DS1. The DS2 recordings were purposely captured in a free-ranging environment; thus, recordings included acoustic signals corrupted by typical external noises, such as blowing wind, cow calls, cow steps, vehicle engines and human voices, among others, and were used to evaluate the proposed algorithm under adverse acoustic conditions.

All segments were labelled aurally by two experts in cows foraging behaviour and with prior experience to identify and classify individual JM-events associated with grazing and rumination. The supervised labelling was done by one of the experts, with results double inspected and checked by the second expert. In most of the cases experts largely agreed with the labelling process, but when there was disagreement, both experts worked together to convey final JM-events classification decisions. The same procedure was applied successfully in Chelotti et al. (2018, 2016, 2020); Vanrell et al. (2018). This supervised JM-event labelling was used as control reference for the purposes of comparing and testing the performance of the algorithm, and to evaluate the quality of sound signals.

The noise effect of environmental sounds on grazing and rumination signals was measured using the JM modulation index (MI_{JM}) and the SNR. The MI_{JM} denotes the suitability of signals for recognition of JM-events in a given segment and it is defined as the ratio between the mean sound intensity during JM-events to the mean sound intensity during the pauses inter-JM-events. The SNR is used to indicate how the proper JM-events classification is affected by noise. A detailed description of the computation of quality factors is given in Appendix A. Both factors were computed for each segment and then averaged per foraging activity (rumination or grazing) in each dataset. Reported values are given in Table 1.

Table 1: Average value for the noise effect on acoustic signals of foraging activities collected in the DS1 and DS2 datasets

	Rumination		Grazing	
	DS1	DS2	DS1	DS2
$\mathrm{MI}_{JM}(\%)$	37.9	21.5	78.1	45.8
SNR (dB)	5.5	2.7	11.7	9.8

2.5. Performance measures

To evaluate the algorithm, comparisons between JM-events recognised by the algorithm vs. JM-events classified aurally were mutually synchronised and compared. This temporal synchronisation and comparison was performed using the HResults tool, which is available as part of the HTK speech analysis toolkit software (HTK 3.4.1, Cambridge University, UK). JM-event synchronisation, matching algorithm recognised and aurally classified JM-event sequences was conducted by an optimal string match using a dynamic programming-based string alignment procedure (Young et al., 2002). The recognition statistics outputs were used to evaluate the detection and classification performance metrics.

The JM-event detection only considers the existence or not of an JM-event, ignoring their corresponding class. The detection performance is collectively affected by the number of correct JM-events detected (true positives), the number of undetected JM-events (false negatives) and the number of incorrect JM-events detected (false positives). The effectiveness of the JM-event detector is reported through metrics for $precision$, $recall$ and $F1 - score$.

The JM-event classification considers the type of JM-events detected. The correct number of previously detected JM-events were used to compute the number of true positives (tp_i), the number of true negatives (tn_i), the number of false positives (fp_i), and the number of false negatives (fn_i) JM-events for each JM-event class i, respectively. Classification performances were averaged by JM-event class (macro-averaging) and reported as F1-score ($F1-score_M$), arithmetic precision ($precision_M$) and arithmetic recall ($recall_M$), (Sokolova and Lapalme, 2009), and their equivalent geometric precision ($precision_G$) and geometric recall ($recall_G$) (Ballabio et al., 2018). The dispersion of the classification performance per JM-event class tends to affect more negatively geometric averages than arithmetic averages.

2.6. Experimental setup

The proposed CBEBA was coded and the experiments were both carried out using Matlab R2019b (MathWorks, Natick, MA, USA) in a personal computer with an Intel Core i7-4790 3.6 GHz with 16 GB of RAM. The MLP network classifier was implemented using the Levenberg-Marquardt backpropagation algorithm optimising the mean square error (Demuth et al., 2014). Inputs were normalised and output labels binarised to match MLP output. MLPs classifiers with five input neurons (number of input features), four output neurons (number of output labels corresponding to JM-events) and one hidden layer were used. The hyperparameters correspond to the learning rate (from 0.1 to 0.001) and numbers of neurons in the hidden layer (from three to seven) were optimised using the grid search method.

Segments in each dataset were split into two parts. The first part included one-third of the segments and served to tune the internal parameters of the algorithm. The second part included the remaining two-thirds of the segments and were used to validate the algorithm (Fig. 3).

In the first part, a train/test split scheme was performed to tune the internal parameters of CBEBA using a grid search method. Four grazing segments and four rumination segments were used for training the model (light red background segments in Fig. 3). A MLP was trained for each internal parameters configuration optimising its hyperparameters. Additional synthetic JM-events for the minority class (bites) were generated using the adaptive synthetic (ADASYN) algorithm. This step was applied to control for JM-events class imbalances during the algorithm training phase (He et al., 2008). Two grazing segments and two rumination segments were used for evaluating the performance of the system for each configuration (dark red background segments in Fig. 3). The optimal internal parameters configuration that maximised the condition $F1-score_M$ was founded varying the following parameters: (i) lower (f_{lower}) and (ii) upper (f_{upper}) frequencies during pre-processing stage, (iii) the cut-off frequency (f_{cut}) involved during envelope computation, and (iv) the envelope sub-sampling frequency (S_s). The corresponding optimal values for both datasets were f_{lower} = 175 Hz,

$f_{upper} = 900$ Hz, $f_{cut} = 9$ Hz and $S_s = 150$ Hz. The variation range of the parameters and their relationships with JM-event features computations are available in the supplementary material.

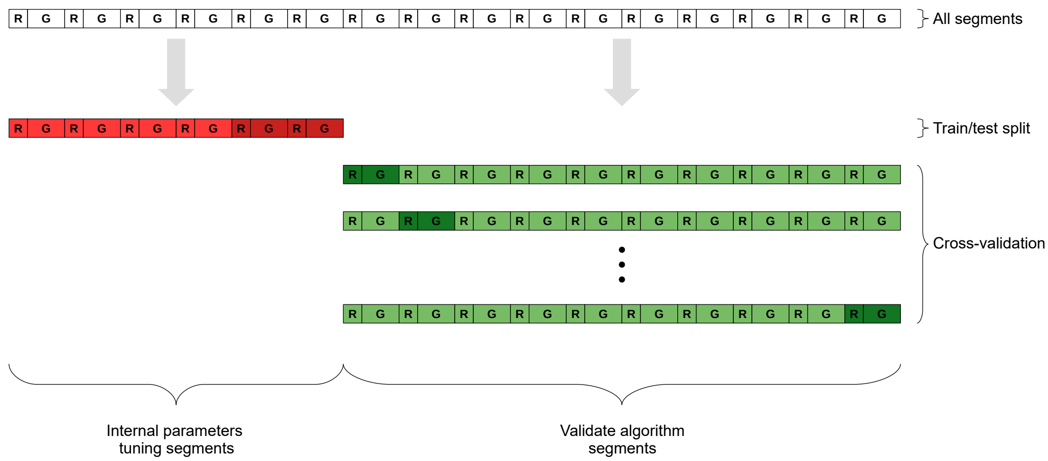

Figure 3: Configuration of the segment fragmentation used to perform the experiments in each dataset. R labels correspond to 75 s *rumination-segments* and G labels correspond to 150 s *grazing-segments*.

The optimal internal parameters obtained in the first part of the split dataset were used in the second part for the final validation of the algorithm in a cross-validation scheme with the corresponding part of the data (Fig. 3). Since grazing and rumination segments are related to different JM-events class, in each iteration, one grazing segment and one rumination segment were selected for the purpose of testing (dark green background segments in Fig. 3), while the remaining segments were used to train the model (light green background segments in Fig. 3). The two segments chosen for testing were only used once and they were not used for testing in subsequent iterations. During the algorithm training phase, synthetic bites were generated with the ADASYN algorithm (He et al., 2008). Performance metrics informed were obtained averaging all iterations. Cross-validation was instrumented as a comprehensive statistical method for evaluation of CBEBA.

Evaluations were conducted according to the following four dataset combinations: i) training with DS1 and testing with DS1 (referred as DS11); ii) training with DS1 and testing with DS2 (referred as DS12); iii) training with DS2 and testing with DS1 (referred as DS21); and iv) training with DS2 and testing with DS2 (referred as DS22).

3. Results

3.1. Qualitative results

A qualitative analysis of the effect on JM-event class separation of the proposed features was performed using a t-distributed stochastic neighbor embedding (t-SNE) analysis (van der Maaten and Hinton, 2008). Fig. 4 shows the result of the dimensionality reduction analysis applied to both sets of JM-events features proposed in this work (Fig. 4a and Fig. 4b) and those sets of features previously applied in the former CBIA (Fig. 4c and Fig. 4d). These figures show the separation of JM-events in four classes for the proposed algorithm (B, CB, RC and GC) and three for CBIA (B, CB and C, where RC and GC are joined as C). Fig. 4a and Fig. 4c show the results for both algorithms in noiseless conditions using the DS1 dataset. B and CB are grouped into two detached clusters with few mixed data points between them. C are also grouped into two clear clusters for both sets of features. In the CBIA case (Fig. 4c) both clusters correspond to the same JM-event class, while in the CBEBA case (Fig. 4a) each cluster corresponds to a different JM-event class. The clusters obtained with the proposed set of features are detached with few mixed data points between them (Fig. 4a). RC and GC are clearly separated, which supports the hypothesis for differences in acoustic properties associated with chewing of fresh (GC) and preprocessed (RC) material. Only B are mixed with both C and CB. CB mixed with GC are negligible. On the other hand, the clusters obtained with the CBIA set of features are detached for some groups and they have mixed data points between them (Fig. 4c). The C cluster is blended with the B cluster. CB form two separated clusters with one of clusters slightly mixed with C events. C form two isolated clusters, which may be explained by the presence of GC and RC (Fig. 4c).

Fig. 4b and Fig. 4d show the results in noisy conditions using the DS2 dataset. Similar separation and characteristics of clusters occurs for the proposed set of features with an increment in clusters overlapping (Fig. 4b). For the CBIA set of features, the clusters are less dense with a considerable number of mixed data points. In this sense, C are grouped in two clusters. One of them presents low density and is mixed with data points corresponding to B and CB. The cluster corresponding to B is low density grouped and mixed with C and CB points. A similar situation occurs with the cluster of CB (Fig. 4d).

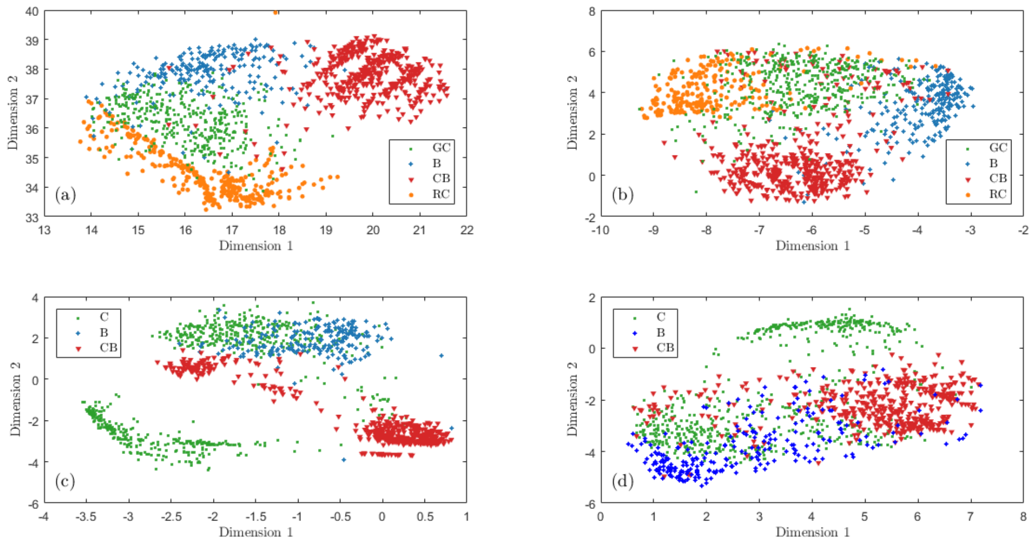

Figure 4: t-SNE analysis of the set of JM-events features used in the proposed CBEBA (panels a and b) and those used in previous CBIA (panels c and d) performed against DS1 noiseless dataset (panels a and c) or DS2 noisy dataset (panels b and d).

3.2. *Algorithm performance metrics*

The proposed CBEBA was evaluated in each of the four possible dataset combinations, and their performance metrics are summarised in Table 2. The detection reached an excellent performance as shown by a *recall* of ∼98.0% in all dataset combinations and a *precision* of ∼99.0% for DS11 and DS21. However, *precision* decreased to values of ∼96.0% for DS12 and DS22. A similar behaviour was seen for $F1 - score$. The classification performance was the highest and more consistent for DS11 as shown by all classification metrics with higher mean and lower standard deviation. Likewise, classifications for DS22 showed smaller deviations in the mean value (<2%) and standard deviations (<3%) than DS11. Conversely, DS12 had lower overall classification performance than DS11. This deterioration was relatively small (<4%) for $recall_M$, $recall_G$ and $F1 - score_M$; and greater (>5%) for $precision_M$ and $precision_G$, respectively. The standard deviation for the classification metrics of DS12 doubled the standard deviation observed for DS11 due to the influence of environmental noise. The classification performance for DS21 had lower $recall_M$ (-3.8%), $F1 - score_M$ (-1.5%) and $recall_G$ (-6.5%), with slighter improvements for $precision_M$ (0.9%) and $precision_G$ (1.2%) than those corresponding to DS22. The standard deviation for the

Table 2: Comparative metrics for CBEBA detection and classification performance (mean ± SD; %) for different combinations of training and testing dataset variants.

		DS11	DS12	DS21	DS22
Detection	$Recall$	98,4 ± 1,0	98,0 ± 0,9	97,9 ± 1,3	98,4 ± 1,2
	$Precision$	99,2 ± 0,5	95,8 ± 2,1	98,7 ± 0,8	96,1 ± 2,6
	$F1-score$	98,8 ± 0,6	96,9 ± 1,4	98,3 ± 0,9	97,2 ± 1,8
Classification	$Recall_M$	92,0 ± 2,6	89,8 ± 4,3	87,5 ± 3,8	91,3 ± 3,8
	$Precision_M$	92,2 ± 2,8	87,0 ± 5,5	91,6 ± 3,1	90,7 ± 5,1
	$F1-score_M$	92,1 ± 2,4	88,4 ± 4,7	89,5 ± 3,2	91,0 ± 4,2
	$Recall_G$	88,9 ± 3,7	87,9 ± 4,8	81,8 ± 5,7	88,3 ± 4,3
	$Precision_G$	89,3 ± 4,3	81,5 ± 8,6	88,9 ± 4,7	87,7 ± 6,6

classification metrics of DS21 were smaller than those for DS22 due to reductions of environmental noise associated with the sound signal of DS1.

Figure 5: Confusion matrices for classification of grazing-chews (GC), bites (B), chew-bites (CB) and rumination-chews (RC) by CBEBA in the a) DS11, b) DS12, c) DS21, and d) DS22 dataset combinations, respectively.

The recognition results across the four dataset combinations, obtained by accumulating the results of each segment used for testing, are presented in the confusion matrices in Fig. 5. Each row represents the distribution of true JM-events over the event class into which they were classified. The rumination-chew events were almost perfectly classified. The grazing-chew and chew-bite events reached recognition rates higher than 86.2% in all dataset combinations. Grazing-chews and chew-bites were better classified in DS1 (DS11 and DS21) than in DS2 (DS12 and DS22). There was lower confusion between datasets in DS1 (3.8% GC classified as CB and 2.1% CB classified as GC in DS11, and 1.7% GC classified as CB and 5.6% CB classified as GC in DS21) than in DS2 (8.2% GC classified as CB and 6.4% CB classified as GC in DS12, and 4.9% GC classified as CB and 8.8% GC classified as GC in the DS22). Particularly, GC and B were the most affected JM-events in mismatch conditions. The recognition rate of GC increased, whereas the recognition rate of B decreased when CBEBA was trained using the noisy dataset (DS2).

3.3. Comparisons between CBEBA and CBIA

A direct comparison between the CBEBA and the former CBIA is not possible because both algorithms are based on a different classification of JM-events. Therefore, one alternative is to combine the grazing and rumination chews in CBEBA into a single chew (C) class in order to have the same type and number of JM-event classes as in CBIA (C, B and CB classes, respectively).

The comparative performance for detection and classification of JM-events between CBIA and CBEBA are presented in Table 3. The CBIA was configured, trained and tested in the same way as the CBEBA. Performance metrics discrepancy between CBEBA and CBIA were evaluated to be statistically significant ($p < 0.05$), using a Wilcoxon signed-rank test (Wilcoxon, 1945).

CBEBA showed better overall performances than the former CBIA. Both algorithms showed similar *recall* for detection in all dataset combinations ($p > 0.05$). However, the *precision* and $F1-score$ indicated a better overall JM-event detection performance for CBEBA than CBIA, respectively ($p < 0.05$). In noisy testing environments (DS12 and DS22), CBEBA achieved better metrics for *precision* ($\sim$96% vs. $\sim$92%) and $F1-score$ ($\sim$97% vs. $\sim$95%) for detection than CBIA. This discrepancy between metrics is likely due to the apparent effects of noises on the *precision* and

Table 3: Comparative performance metrics (mean ± SD; %) for CBIA and CBEBA across the four dataset combinations. Metric values with green background within same rows and dataset combinations differ significantly, whereas metric values with pink background within same rows and dataset combinations show no significant difference ($p < 0.05$; Wilcoxon signed-rank test).

		DS11		DS12		DS21		DS22	
		CBIA	CBEBA	CBIA	CBEBA	CBIA	CBEBA	CBIA	CBEBA
Detection	*Recall*	**98,8 ± 0.7**	98.4 ± 1.0	**98.5 ± 0.7**	98.0 ± 0.9	**98.0 ± 0.9**	97.9 ± 1.3	**98.7 ± 0.6**	98.4 ± 1.2
	Precision	98.8 ± 0.8	**99.2 ± 0.5**	91.7 ± 2.8	**95.8 ± 2.1**	98.0 ± 0.7	**98.7 ± 0.8**	91.9 ± 2.6	**96.1 ± 2.6**
	$F1-score$	**98.8 ± 0.5**	98.8 ± 0.6	95.0 ± 1.7	**96.9 ± 1.4**	97.8 ± 0.5	**98.3 ± 0.9**	95.1 ± 1.5	**97.2 ± 1.8**
Classification	$Recall_M$	91.7 ± 3.3	**91.8 ± 3.4**	72.9 ± 5.0	**90.5 ± 4.6**	82.0 ± 6.8	**92.9 ± 2.6**	86.7 ± 6.2	**90.0 ± 4.2**
	$Precision_M$	**93.7 ± 2.4**	92.1 ± 3.5	79.6 ± 6.7	**85.9 ± 5.7**	83.0 ± 4.9	**85.1 ± 4.8**	86.8 ± 5.5	**90.2 ± 5.4**
	$F1-score_M$	**92.5 ± 1.8**	91.9 ± 2.4	77.5 ± 5.2	**88.0 ± 4.4**	82.9 ± 5.1	**88.8 ± 3.1**	87.3 ± 5.0	**90.0 ± 4.3**
	$Recall_G$	90.8 ± 3.9	**91.5 ± 3.8**	68.0 ± 7.6	**90.2 ± 5.1**	80.2 ± 7.8	**92.7 ± 2.8**	85.3 ± 8.2	**89.5 ± 4.7**
	$Precision_G$	**93.5 ± 2.8**	91.6 ± 4.0	78.5 ± 7.6	**84.3 ± 7.3**	81.6 ± 5.8	**83.3 ± 5.7**	85.7 ± 7.3	**89.6 ± 6.0**

$F1-score$ when each algorithm was trained with a noiseless dataset. This fact is noteworthy for DS12 compared with DS11, where the performance deterioration for CBEBA was roughly half of that observed for CBIA.

The JM-event classification showed similar results. Both algorithms had similar performance metrics for the noiseless train/test dataset combination (DS11), but CBEBA had better classification performance (high mean and low standard deviation of metrics) than CBIA in the presence of noises. Compared to DS11, the performance metrics for JM-event classification for DS12 declined, but the decline for CBEBA was less than half of that observed for CBIA. For the cases of noisy training conditions (DS21 and DS22), CBEBA achieved better overall metrics for JM-event classification performance than CBIA, respectively, indicating a greater sensitivity of CBIA to environmental noise (Table 1). In addition, the difference between $recall_M$ and $recall_G$ was smaller for CBEBA than CBIA in the same dataset combination, supporting greater dispersion in the recall per JM-event class for CBIA than CBEBA. The same situation occurred for $precision_M$ and $precision_G$ in DS22, further supporting the proposed set of JM-events features for robust detection and classification of broad sets of JM-event classes using CBEBA.

The confusion matrices in Fig. 6 and Fig. 7 show the JM-events recognition rates obtained by CBEBA and CBIA. Fig. 6 shows comparative results for JM-events classification in noiseless data (DS11) and noisy data (DS12) obtained using noiseless data for training. A similar situation obtained using noisy data for training (DS21 and DS22) is shown in Fig. 7. The first row in figures (Fig. 6a and Fig. 6b, and Fig. 7a and Fig. 7b) show the results for CBEBA, whereas the second row in figures (Fig. 6c and Fig. 6d, and Fig. 7c and Fig. 7d) show the results for CBIA.

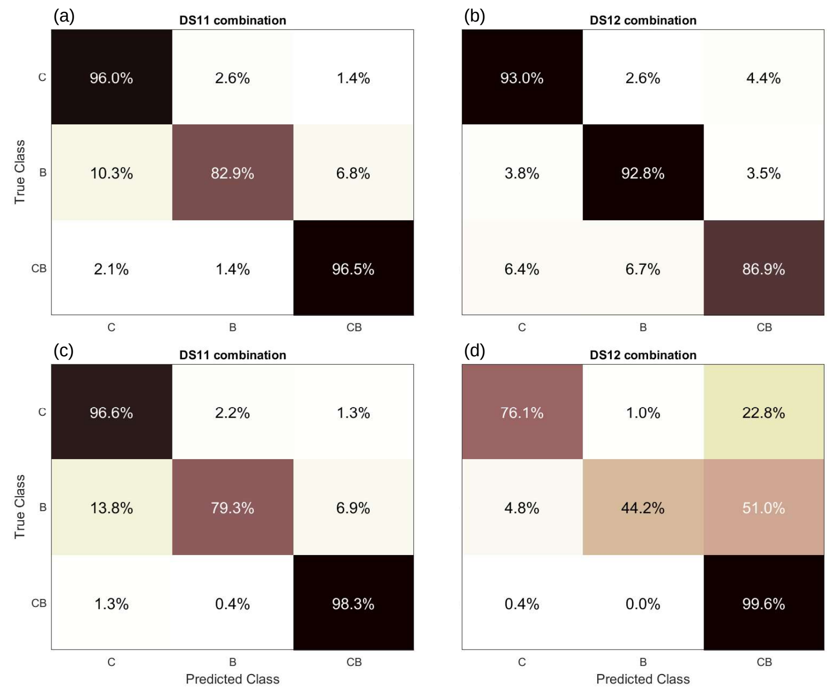

Figure 6: Confusion matrices of the DS11 and DS12 dataset combinations showing recognition rates for chew (C), bite (B) and chew-bite (CB) events classified respectively by the CBEBA (panels a and b) and CBIA (panels c and d) algorithms.

The confusion matrices for DS11 in Fig. 6 showed a slight improvement in the recognition of bites for CBEBA vs. CBIA. Compared to findings shown for CBEBA in DS11, confusion matrices for CBEBA in DS12 showed a similar average JM-events recognition rate (91.8% vs. 90.9%, respectively) with a 9.6% deterioration for recognition of chew-bites and 9.9% improvement for recognition of bites. Compared to results for CBIA in DS11, findings for CBIA in DS12 showed a clear decline in average JM-events recognition rate (91.4% vs. 73.3%, respectively), due to a 20.7% decline in recognition of chews, a 35.1% decline in recognition of bites, and only a 1.3% improvement in recognition of chews-bites as a consequence of the great confusion of chews and bites with chew-bites.

The confusion matrices for DS21 in Fig. 7 showed a slightly better average JM-events recognition rate for CBEBA (84.9%) than CBIA (83.2%). The confusion matrix for DS21 showed that CBEBA correctly recognised more

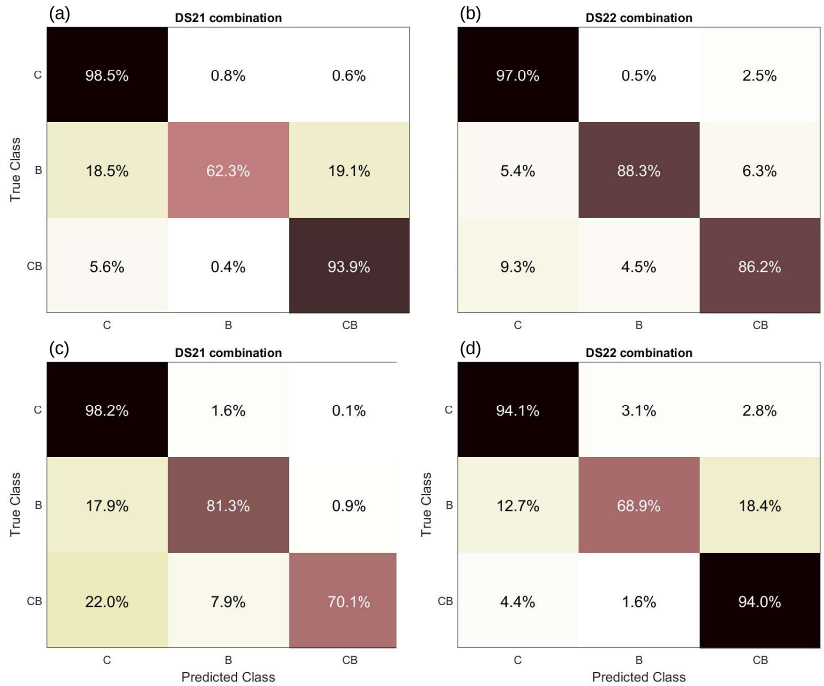

Figure 7: Confusion matrices of the DS21 and DS22 dataset combinations showing recognition rates for chew (C), bite (B) and chew-bite (CB) events classified respectively by the CBEBA (panels a and b) and CBIA (panels c and d) algorithms.

than 93.9% of combined grazing and rumination chews and chew-bites, but only 62.3% of bites. Conversely, CBIA achieved a better recognition of bites (81.3%) at the expense of a lower recognition of chew-bites (70.1%). Furthermore, CBIA presented a much lower recognition of bites in DS22 than CBEBA (68.9% vs. 88.3%, respectively). However, it is noteworthy that regarding DS22 the average JM-events recognition rate for CBEBA declined 5.6% compared to a decline of 2.4% for CBIA in DS21, with a 26% decline in the recognition of bites and despite a 1.5% and 7.7% improvement in the recognition of chews and chew-bites for CBEBA, respectively. Compared to results for CBIA in DS22, findings for CBIA in DS21 showed lower average JM-events recognition rate (85.6% vs. 83.2%, respectively) due to a 23.9% decline in the recognition of chew-bites, and despite the 4.1% and 12.4% improvement in the recognition of chews and bites, respectively.

Table 4 shown the computational cost of CBEBA and CBIA, both ran on a similar MLP neural network for JM-event classification, expressed in terms of the number of operations per second (ops/s). A detailed description of the operations and assumptions is given in Appendix B. The overall computational cost of CBEBA was 11% higher than that for CBIA, mainly due to the additional costs associated with the signal conditioning, JM-event classification and updating of internal algorithm parameter tasks.

Table 4: Partitioned computational costs for the CBEBA and CBIA expressed in operations per second.

Algorithm computational stage	CBIA [1] [ops/s]	CBEBA[2] [ops/s]
Signal conditioning and pre-processing	10000	16000
Event detection and feature extraction	28800	26110
Event classification	160	690
Internal parameters update	-	378
Global cost	38960	43178

[1] Considering k=2 kHz, n=100 Hz and a MLP classifier.
[2] Considering k=2 kHz, n=150 Hz and a MLP classifier.

4. Discussion

Previous deployment of acoustic telemetry to monitor livestock foraging behaviour showed a promising future for the technique, but a major bottleneck has been the need for more appropriate and robust analytical procedures to deal with issues associated with environmental noise. The present study builds upon a previous methodology (CBIA) to analyse JM of grazing cattle and incorporates analytics of novel sound signal features combined with machine learning techniques for a better detection and classification of JM-events in noisy environments.

A major advantage for the proposed CBEBA has been the expansion in detection and classification of JM-event classes. While previous attempts (Chelotti et al., 2018, 2016; Deniz et al., 2017; Milone et al., 2012) did not differentiated chews among those associated with ingestive grazing and those for rumination, a qualitative analysis of the proposed and previous (CBIA) set of JM-events features showed that chew events are always clustered into two groups, independently if we explicitly differentiate chews associated with ingestive grazing and rumination (the proposed set of features) or not (the CBIA set of features) (Fig. 4). This organisation of the data indicates that this desegregation of the chew events cluster is possible and could be proposed to improve the recognition capabilities of the algorithms. Thus, the present CBEBA method successfully classified JM-events into exclusive bites, grazing-chews, composite chew-bites and exclusive rumination-chews. Further, the computation of the instantaneous signal power for determination of JM-event energy as a new discriminative feature allowed a better discrimination of chews associated with ingestive grazing and rumination. A plausible explanation could be related to differences in the water content of grazed and ruminated forages. Moisture content can be classified into internal or external, depending on whether it belongs or not to cell wall contents. Likewise, as masticatory sounds produced by chewing are partly due to a rupture of cell walls and extrusion of internal water and contents, we hypothesise that grazing-chews may have produced sound JM-events of greater total energy than rumination chews because the water and contents of intact cells in the ingested forage material were greater than that for the partially processed and degraded material regurgitation during rumination (Galli et al., 2020).

The suitability and comparative ability of CBEBA for detection and classification of JM-events were evaluated using the same and different dataset combinations for the algorithm training and testing. This approach involved datasets acquired in a noisy and noiseless environment, thus allowing for direct evaluating of noise effects and the sensibility to mismatch condition. CBEBA presents good overall results for detection and classification of JM-events. In all training and testing dataset combinations, including noisy datasets, the *recall* for detection was close to 98%, which indicates a low rate of undetected JM-events. The *precision* for detection reached slightly better values for test with noiseless data (DS11 and DS21) than for test with noisy data (DS12 and DS22) ($\sim$99% vs. $\sim$96%), which was associated with a greater number of false JM-events detected when testing under noisy scenarios. This finding supports promising applications of CBEBA for deployment in noisy environments. Furthermore, all classification metrics for DS22 decreased less than 1.6% compared to the best noiseless scenario (DS11) (Table 3) and a recognition rate higher than 82% for all JM-events (with the particularity that rumination-chews were almost perfectly classified) was observed in DS11 and DS22 (Fig. 5), suggesting that the features choices are robust in a noisy environment.

The CBEBA performance using different training data was evaluated. Better results in all classification metrics were observed when CBEBA, tested with noiseless data, was trained using noiseless data (DS11) than those using noisy data (DS21). Further, similar JM recognition rates of grazing-chews, chew-bites and rumination-chews were observed, but the JM recognition rate of bites decreased considerably to 62.3% in DS21 (Fig. 5). This suggests that the additional information provided to algorithm training by noise had a negative influence on noiseless JM-events classifications success. On the other hand, CBEBA, tested with noisy data, showed a decrease in all classification metrics when it was trained using noiseless data (DS12) compared to those using noisy data (DS22). A plausible explanation could be related to the fact that the noise provided additional useful information during the algorithm training for correct noisy JM-events classifications.

Additionally, changing either the sampling frequency of the input audio (fulfilling the Nyquist–Shannon sampling theorem) or increasing the word-length of the digitised input audio above eight bits had a negligible variation (<0.5%) in all performance metrics for CBEBA, and consistent with previous work by Chelotti et al. (2016). The computational cost of CBEBA was 11% higher than CBIA (43178 ops/s vs 38960 ops/s). Thus, the low computational cost, the low input sampling frequency, and the reduced input word-length representation required by CBEBA make it suitable for use as part of portable components for embedded devices. Microcontroller-based embedded systems can reduce their power consumption working at low clock frequency. This fact allows the devices to operate for extended periods of time. In the case that CBEBA would like to be implemented in a microcontroller to operate in real-time, the amount of processing operations associated to registers and memory access would be taken into account. If it is considered that these operations are approximately three times higher than the previously 43178 arithmetic and logic operations, the total number of processing operations for the microcontroller would be 172,712 ops/s. Furthermore, the execution time depends on the architecture and the operating frequency of the microcontroller. For example, using an ARM Cortex-M4 microcontroller (ATSAM4LS2A, Microchip Technology Inc., Chandler, AZ, USA) operating at 1 MHz and considering that performing an operation may require up to 4 clock cycles, CBEBA would require approximately 173 ms. In the case of CBIA, the execution time would be approximately 156 ms following the same analysis. The idle time available until the next JM-event could be used to perform peripheral management or enter in a low power mode.

Accurate measurement of herbage intake is critical to advance knowledge on foraging mechanisms that can result in a more efficient production and utilization of herbage resources along with a more profitable and affordable animal production. According to pioneer acoustic works (Laca et al., 2000) and earlier applications of the technique with sheep (Galli et al., 2011) and dairy cattle (Galli et al., 2018), the sound energy enclosed in masticatory (chewing and biting) events serves as robust predictor of herbage intake and could be scaled successfully to monitor changes in animal production, health or fitness, long-term. Therefore, the automated computation of sound energy as a novel analytical property in CBEBA is highly relevant and very appealing. This new property suggests a new opportunity to automate CBEBA to routinely monitor and assess use livestock foraging behaviour to manage feed intake (Galli et al., 2018).

5. Conclusions

In this study, the CBEBA was tested as robust alternative for acoustic detection and classification of JM of grazing cows, typically involving noisy environments. The algorithm integrated six generic stages and incorporated computations based on the instantaneous power signal rather than on the intensity signal to improve the JM-event detection sensibility. A novel and extended set of robust features are extracted and combined using machine learning techniques to recognise JM-events. Specific performance metrics for pattern recognition were used to evaluate the results of CBEBA in noisy, noiseless, and crossed scenarios. Overall, CBEBA showed good results, indicating adequacy and scalability of the algorithm for acoustic analytics in broad sets of acoustic scenarios.

Specifically, notable results were obtained in noisy and crossed algorithm training and testing scenarios outperforming the algorithm from the state-of-the-art. To the best of our knowledge this is the first algorithm that is able to distinguish chew events associated with ingestive and rumination activities in grazing cows while dealing accordingly with augmented noises typical of grazing environments. The increase in the computational cost of CBEBA compared to CBIA remained in the same order of magnitude, suggesting feasibility for efficient implementation in a low-cost embedded device for real-time operation.

Future work must include evaluations of the algorithm in more challenging animal husbandry scenarios such as dairy cattle barns and beef cattle feed yards exhibiting louder and more varied array noises as plausible sources of signal interference and degradation. Likewise, future work is needed to scale up uses of CBEBA as part of low cost online applications to monitor animal activities with regard to feed intake, health, breeding, parturition or production. Finally, the recognition performance of the algorithm could be improved by including more complex processing techniques such as using a deep learning approach combining the JM features extraction and JM classification at the expense of an increased computational cost.

Acknowledgments

This work has been funded by Universidad Nacional del Litoral, PACT CAID 2011–525, Universidad Nacional de Rosario, projects 2013-AGR216, 2016-AGR266 and 80020180300053UR, Agencia Santafesina de Ciencia, Tecnología e Innovación (ASACTEI), project IO-2018–00082, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), project 2017-PUE-sinc(i). Also, this work was funded under the USDA-NIFA MICL0222 and MICL0406 projects, and direct support from AgBioResearch-MSU. Authors would like to thank the dedication and perceptive help by the KBS Robotic Dairy Farm staff, and Campo Experimental J. Villarino Dairy Farm staff for their assistance and support during the completion of this study.

AUTHORSHIP STATEMENT

LSMR, JRG, LLG, SAU, HLR and AMP participated in conceptualization; LSMR participated in software stage; LSMR, JOC, JRG and AMP participated in the data curation; LSMR and JOC participated in the formal analysis; LSMR and JOC participated in investigation stage; LSMR, JOC and SRV participated in methodology, validation and visualization stages; JRG, LLG, SAU, SRV and HLR participated in the funding acquisition; LSMR, LLG and HLR participated in project administration; LSMR, JOC, SRV, SAU and LLG contributed to the writing and reviewing of the original draft; All the authors reviewed and approved the manuscript.

Appendix A. Signal quality factors

The quality of the signals available in the two datasets depends on both, the JM and the noise levels. The noise level in corrupted signals can affect the recognition and classification performance.

Given that JM during rumination and grazing feeding activities are performed rhythmically, the variations between the mean sound intensity during

JM-events x_E and the mean sound intensity inter-JM-events x_{IE} are used to indicate the suitability for recognition of true JM-events in a given segment by the following JM modulation index $MI_{JM} = (\overline{x_E} - \overline{x_{IE}})/(\overline{x_E} + \overline{x_{IE}})$, with:

$$\overline{x_E} = \frac{1}{l_s} \sum_{k=1}^{l} x^2[k] \cdot w[k] \tag{A.1}$$

$$\overline{x_{IE}} = \frac{1}{l_n} \sum_{k=1}^{l} x^2[k] \cdot (1 - w[k]) \tag{A.2}$$

where $x[k]$ is the corrupted segment, l is the length of a segment in samples, l_s and l_n are the number of samples with and without JM activity, respectively, and $w[k]$ is a logical function indicating the presence of an JM-event in the k-th sample.

Regarding the classification task, the noise present during JM-events must be isolated. To estimate a noise-free activity signal $\widehat{s}[k]$ and a noisy signal $\widehat{n}[k]$, a multiband spectral subtraction algorithm assuming uncorrelated additive noise in signal segments was used (Loizou, 2013). The criterion used to indicate the difficulty to classify JM-events in a given segment is the SNR, which is computed as follow:

$$SNR(dB) = 10log\left(\sum_{k=1}^{l} \widehat{s}^2[k]\right) - 10log\left(\sum_{k=1}^{l} \widehat{n}^2[k]\right) \tag{A.3}$$

Appendix B. Computational Cost

The computational cost of the CBEBA stages depends on the selected classifier, the sampling frequency of the input signal k, and the sub-sampling frequency n, which also determines the size of buffers. In order to obtain a simple comparison with other online algorithms, a MLP classifier using two JM-events per second was considered in this analysis. Worst-case scenarios were considered for each stage in order to get a theoretical upper bound. The required number of operations per stage of computation for CBEBA was:

1. Signal pre-processing
   (a) *Bandwidth limitation*: A second-order band-pass filter is applied to avoid undesired noises, which involves 4 multiplications and 3 sums per sample ($7k$ ops/s).

   (b) *Power computation*: 1 multiplication per sample is required to obtain the instantaneous power signal (k ops/s).
2. Buffering
   (a) *Envelope computation*: A second-order IIR low-pass filter is applied to the instantaneous power signal to compute the envelope, which requires 5 multiplications and 4 sums per sample ($9k$ ops/s). In order to reduce the computational cost of the next stages the envelope is decimated, so a counter and a comparison per sample is needed ($2k$ ops/s). These values are stored in $e(n)$ at a rate n (n ops/s).
   (b) *Frame-energy computation*: An internal accumulator computes the sum of the energy signal (k ops/s). At a rate n the accumulator value is stored in $f(n)$ and it is reseted ($2n$ ops/s).
3. JM-event detection
   (a) *Possible event*: Finding the first valid peak of the envelope requires 1 subtraction and 1 comparison per JM-event. Searching for a second peak in the envelope requires $0.275n$ subtractions and $0.275n$ comparisons per JM-event. Centring the position between both peaks requires 1 subtraction and 1 comparison per JM-event. Additionally, updating the threshold $T(n)$ requires $0.375n$ operations per JM-event.
   (b) *Event bouts*: Determining the threshold energy level involves $0.5n+4$ operations per JM-event. Assuming the worst case scenario in which the start and end positions of the candidate JM-event coincide with the start and end positions of $f(n)$, $0.5n$ comparisons and 8 accumulations per JM-event are required.
4. Feature extraction: In this scenario, $0.5n - 1$ sums per JM-event are needed to compute the total energy feature. $0.5n$ subtractions and $0.5n$ comparisons per JM-event are required to determine the sign of the envelope slope feature. Computing the duration of the candidate JM-event requires 1 subtraction per JM-event. The symmetry feature involves a maximum of $0.5n - 1$ comparisons, 1 subtraction and 1 division per JM-event. Finally, the accumulated envelope speed feature requires $1.5n - 1$ operations per JM-event.
5. JM-event classification
   (a) *Event decision*: A candidate JM-event must meet the two conditions previously established. The first condition requires 1 comparison per JM-event. The second condition requires $n - 1$ sums

per JM-event to compute the total energy in $f(n)$, and 1 multiplication and 1 comparisons per JM-event to evaluate the logic condition.

(b) *Event classifier*: According to the possible values of the optimal hyperparameters, the most expensive computational cost is considered. The required operations in the MLP classifiers depend on the number of input neurons (5), the number of hidden layers (1), the count of neurons in the hidden layer (7) and the number of output neurons (4). The cost of these classifiers is 192 operations per JM-event.

6. Threshold algorithms and parameters tuning: Update and compute all values related to $T(n)$ requires $n + 24$ operations per JM-event. Similarly, the parameters related to T_F involve 15 operations per JM-event.

The required total number of operations per second is:

$$f_{CBEBA,MLP}(k, n) = 20k + 17.85n + 500$$

References

Andriamandroso, A., Bindelle, J., Mercatoris, B., Lebeau, F., 2016. A review on the use of sensors to monitor cattle jaw movements and behavior when grazing. Biotechnologie, Agronomie, Société et Environnement 20.

Andriamandroso, A.L.H., Lebeau, F., Beckers, Y., Froidmont, E., Dufrasne, I., Heinesch, B., Dumortier, P., Blanchy, G., Blaise, Y., Bindelle, J., 2017. 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. 139, 126–137.

Ballabio, D., Grisoni, F., Todeschini, R., 2018. Multivariate comparison of classification performance measures. Chemometrics and Intelligent Laboratory Systems 174, 33–44.

Berckmans, D., 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. Rev. Sci. Tech. 33, 189–196.

Chelotti, J.O., Vanrell, S.R., Galli, J.R., Giovanini, L.L., Rufiner, H.L., 2018. A pattern recognition approach for detecting and classifying jaw movements in grazing cattle. Computers and Electronics in Agriculture 145, 83–91.

Chelotti, J.O., Vanrell, S.R., Milone, D.H., Utsumi, S.A., Galli, J.R., Leonardo Rufiner, H., Giovanini, L.L., 2016. A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle. Comput. Electron. Agric. 127, 64–75.

Chelotti, J.O., Vanrell, S.R., Rau, L.S.M., Galli, J.R., Planisich, A.M., Utsumi, S.A., Milone, D.H., Giovanini, L.L., Rufiner, H.L., 2020. An online method for estimating grazing and rumination bouts using acoustic signals in grazing cattle. Computers and Electronics in Agriculture 173, 105443.

Demuth, H.B., Beale, M.H., De Jess, O., Hagan, M.T., 2014. Neural network design. Martin Hagan.

Deniz, N.N., Chelotti, J.O., Galli, J.R., Planisich, A.M., Larripa, M.J., Rufiner, H.L., Giovanini, L.L., 2017. Embedded system for real-time monitoring of foraging behavior of grazing cattle using acoustic signals. Computers and electronics in agriculture 138, 167–174.

DeVries, T.J., Beauchemin, K.A., Dohme, F., Schwartzkopf-Genswein, K.S., 2009. Repeated ruminal acidosis challenges in lactating dairy cows at high and low risk for developing acidosis: Feeding, ruminating, and lying behavior. Journal of Dairy Science 92, 5067–5078.

Fuentes, A., Yoon, S., Park, J., Park, D.S., 2020. Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. Computers and Electronics in Agriculture 177, 105627.

Galli, J.R., Cangiano, C.A., Milone, D.H., Laca, E.A., 2011. Acoustic monitoring of short-term ingestive behavior and intake in grazing sheep. Livestock Science 140, 32–41.

Galli, J.R., Cangiano, C.A., Pece, M.A., Larripa, M.J., Milone, D.H., Utsumi, S.A., Laca, E.A., 2018. Monitoring and assessment of ingestive chewing sounds for prediction of herbage intake rate in grazing cattle. animal 12, 973–982.

Galli, J.R., Milone, D.H., Cangiano, C.A., Martínez, C.E., Laca, E.A., Chelotti, J.O., Rufiner, H.L., 2020. Discriminative power of acoustic features for jaw movement classification in cattle and sheep. Bioacoustics 29, 602–616.

Greenwood, P.L., Paull, D.R., McNally, J., Kalinowski, T., Ebert, D., Little, B., Smith, D.V., Rahman, A., Valencia, P., Ingham, A.B., et al., 2017. Use of sensor-determined behaviours to develop algorithms for pasture intake by individual grazing cattle. Crop and Pasture Science 68, 1091–1099.

He, H., Bai, Y., Garcia, E.A., Li, S., 2008. Adasyn: Adaptive synthetic sampling approach for imbalanced learning, in: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), IEEE. pp. 1322–1328.

Hodgson, J.G., 1990. Grazing management: science into practice. John Wiley & Sons Inc.

Laca, E.A., Ungar, E.D., Seligman, N.G., Ramey, M.R., Demment, M.W., 1992. An integrated methodology for studying short-term grazing behaviour of cattle. Grass and forage science 47, 81–90.

Laca, E.A., WallisDeVries, M.F., et al., 2000. Acoustic measurement of intake and grazing behaviour of cattle. Grass and Forage Science 55, 97–104.

Loizou, P.C., 2013. Speech enhancement: theory and practice. CRC press.

Michie, C., Andonovic, I., Davison, C., Hamilton, A., Tachtatzis, C., Jonsson, N., Duthie, C.A., Bowen, J., Gilroy, M., 2020. The internet of things enhancing animal welfare and farm operational efficiency. Journal of Dairy Research 87, 20–27.

Milone, D.H., Galli, J.R., Cangiano, C.A., Rufiner, H.L., Laca, E.A., 2012. Automatic recognition of ingestive sounds of cattle based on hidden markov models. Comput. Electron. Agric. 87, 51–55.

Milone, D.H., Rufiner, H.L., Galli, J.R., Laca, E., Cangiano, C.A., 2009. Computational method for segmentation and classification of ingestive sounds in sheep. Computers and Electronics in Agriculture 65, 228–237.

Nydegger, F., Gygax, L., Egli, W., 2010. Automatic measurement of rumination and feeding activity using a pressure sensor, in: International Conference on Agricultural Engineering-AgEng 2010: towards environmental technologies, 6-8 September 2010, Clermont-Ferrand, France, pp. 1–8.

Roden, M.S., 1996. Analog and digital communication systems. Prentice Hall.

Saitoh, T., Kato, Y., 2021. Evaluation of wearable cameras for monitoring and analyzing calf behavior: A preliminary study. Animals 11, 2622.

Sauvant, D., 2000. Granulométrie des rations et nutrition du ruminant. Productions animales 13, 99–108.

Schirmann, K., Chapinal, N., Weary, D.M., Heuwieser, W., von Keyserlingk, M.A.G., 2011. Short-term effects of regrouping on behavior of prepartum dairy cows. Journal of Dairy Science 94, 2312 – 2319.

Schirmann, K., Chapinal, N., Weary, D.M., Vickers, L., von Keyserlingk, M.A.G., 2013. Short communication: Rumination and feeding behavior before and after calving in dairy cows. Journal of Dairy Science 96, 7088 – 7092.

Schirmann, K., von Keyserlingk, M.A.G., Weary, D.M., Veira, D.M., Heuwieser, W., 2009. Validation of a system for monitoring rumination in dairy cows. J. Dairy Sci. 92, 6052–6055.

Smith, D., Rahman, A., Bishop-Hurley, G.J., Hills, J., Shahriar, S., Henry, D., Rawnsley, R., 2016. Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems. Comput. Electron. Agric. 131, 40–50.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45, 427–437.

Stone, A.E., 2020. Symposium review: The most important factors affecting adoption of precision dairy monitoring technologies. Journal of Dairy Science 103, 5740–5745.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. Journal of machine learning research 9, 2579–2605.

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

Watt, L.J., Clark, C.E.F., Krebs, G.L., Petzel, C.E., Nielsen, S., Utsumi, S.A., 2015. Differential rumination, intake, and enteric methane production of dairy cows in a pasture-based automatic milking system. Journal of dairy science 98, 7248–7263.

Werner, J., Leso, L., Umstatter, C., Niederhauser, J., Kennedy, E., Geoghegan, A., Shalloo, L., Schick, M., O'Brien, B., 2018. Evaluation of the RumiWatchSystem for measuring grazing behaviour of cows. Journal of Neuroscience Methods 300, 138–146.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1, 80.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., et al., 2002. The htk book. Cambridge university engineering department 3, 12.

Zehner, N., Umstätter, C., Niederhauser, J.J., Schick, M., 2017. 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 136, 31–41.