A pattern recognition approach for detecting and classifying jaw movements in grazing cattle

José O. Chelotti[a], Sebastián R. Vanrell[a], Julio R. Galli[b], Leonardo L. Giovanini[a], H. Leonardo Rufiner[a,c]

[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, Facultad de Ciencias Agrarias, UNR-CONICET, Argentina*
[c] *Laboratorio de Cibernética, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina*

Abstract

Precision livestock farming is a multidisciplinary science that aims to manage individual animals by continuous real-time monitoring their health and welfare. Estimation of forage intake and monitoring the feeding behavior are key activities to evaluate the health and welfare state of animals. Acoustic monitoring is a practical way of performing these tasks, however it is a difficult task because masticatory events (bite, chew and chew-bite) must be detected and classified in real-time from signals acquired in noisy environments. Acoustic-based algorithms have shown promising results, however they were limited by the effects of noises, the simplicity of classification rules, or the computational cost. In this work, a new algorithm called Chew-Bite Intelligent Algorithm (CBIA) is proposed using concepts and tools derived from pattern recognition and machine learning areas. It includes i) a signal conditioning stage to attenuate the effects of noises and trends, ii) a preprocessing stage to reduce the overall computational cost, iii) an improved set of features to characterize jaw-movements, and iv) a machine learning model to improve the discrimination capabilities of the algorithm. Three signal conditioning techniques and six machine learning models are evaluated. The overall performance is assessed on two independent data sets, using metrics like recognition rate, recall, precision and computational cost.

Email address: `jchelotti@sinc.unl.edu.ar` (José O. Chelotti)

The results demonstrate that CBIA achieves a 90% recognition rate with a marginal increment of computational cost. Compared with state-of-the-art algorithms, CBIA improves the recognition rate by 10%, even in difficult scenarios.

Keywords: Dairy cows, precision livestock farming, acoustic monitoring, signal processing, machine learning.

1. Introduction

Nowadays, the management of livestock grazing systems requires accurate measurement of animal feeding behavior to monitor their health and welfare, as well as to improve the efficiency of resource management. In this regard, much effort has been put into finding suitable techniques for monitoring the feeding behavior of ruminants. A long-term analysis of such behavior distinguishes two major activities: rumination and grazing, which last from few minutes to hours. On a short-time scale, these activities are composed by a sequence of three jaw movements: bites, chews and chew-bites (Ungar et al., 2006; Milone et al., 2012). A grazing bite includes the apprehension and severance of forage, while a grazing or rumination chew includes the crushing, grinding and processing of ingested pasture. The chew-bite is another grazing event that results from the overlapping of chew and bite events in the same jaw movement. While the number and characteristics of jaw movements change according to animal and environmental factors, monitoring them can provide useful indicators of animal health, welfare, nutritional status, and feeding activities (grazing and rumination) (De Boever et al., 1990).

Early strategies for monitoring feeding behavior were based on direct observation and in recent times by visualization of video recordings. However, both methodologies are costly and impractical for monitoring large herds (Milone et al., 2009). In last decades, other methods based on pressure sensors, accelerometers and microphones have been studied (Andriamandroso et al., 2016). Most of them focus on recognizing long-term activities (rumination and grazing) rather than individual jaw movements. Detection of jaw movements can be performed with nose-band pressure sensors (Nadin et al., 2012; Zehner et al., 2017) and accelerometers (Tani et al., 2013; Oudshoorn et al., 2013; Andriamandroso et al., 2016) but the available results indicate that classification requires further development to be reliable and automatic. In addition, a separate classification of chews, bites, and chew-bites cannot be done because the compound chew-bite cannot be identified by these methods. By contrast, several studies have shown that acoustic monitoring can overcome these limitations.

An accurate tracking of animal diet can be accomplished by analyzing the sounds related to jaw movements (Alkon and Cohen, 1986; Alkon et al., 1989; Laca et al., 1992; Ungar et al., 2006). Biting and chewing sounds are produced while plant structures are comminuted by jaw movements. The sound is transmitted, filtered and modified by the bones, cavities and soft tissues of the animal's head. It can be recorded and collected in a non-invasive way without affecting the natural behavior of animals (Laca et al., 1992; Klein et al., 1994; Nelson et al., 2005). However, acoustic analysis is a complex task, particularly in noisy environments like cattle barns, and known applications usually show high computational-cost.

Among automatic recognition systems based on sound analysis, just few of them deal with detection and classification of jaw movements problem. Milone et al. (2012) developed an algorithm based on hidden Markov models, that hereafter will be referred as CBHMM (Chew-Bite Hidden Markov Model), for detecting and classifying jaw movements. Using probabilistic models and spectral-domain features it achieves up to 85% recognition rate, but it shows a high computational cost. Navon et al. (2013) implemented an algorithm for event detection that used a machine-learning technique to analyze time-domain features of ingestive sounds. The algorithm achieved up to 94% detection rate. However, the event classification (bite, chew and chew-bite) was not performed in this work.

The development of a recognition system based on analysis of acoustic signals should consider the following situations:

1. Input sound signals are affected by environmental noises, which can degrade the signal-to-noise ratio (SNR) and diminish the overall system performance. For instance, trends (low-frequency time-varying noises) are more intense when cattle stay in barns, where there are more noise sources (e.g. machinery and other animals) that are also intensified by the room reverberation.
2. Recognition of ingestive events is a combination of detection and classification. Former methods showed that detection can be successfully performed with high accuracy whereas classification typically requires more powerful methods to achieve high recognition rates.
3. Precision livestock farming area aims at low-cost algorithms in order to

embed and execute them in real-time within low-performance wearable devices. In this way, the monitoring system could be scaled for its application on large herds.

Recently, Chelotti et al. (2016) proposed an algorithm called Chew-Bite Real-Time Algorithm (CBRTA) for detection and classification of ingestive events. The algorithm used heuristic rules derived from expert knowledge, reaching recognition rates up to 97.4% for detection and 84.0% for classification of events at a low computational cost. However, trends in the input signals negatively affect the event detection because it is based on time-varying thresholds. On the other hand, the simple set of rules proposed may not exploit the whole potential of the features employed. For instance, the shape and duration of a detected event are the only features used to differentiate compound events (chew-bites) from simple events (chews or bites), which is an oversimplified representation of the events, leading to poor recognition rate. Thus, it is desirable to consider machine learning techniques that are able to generate optimal decision regions, improving the classification performance at the expense of reasonable increments of computational cost (up to 50% of CBRTA cost), which will imply neglectable effects on the algorithm's computational requirements due to the extremely low computational cost of the CBRTA.

In this paper, a new algorithm called Chew-Bite Intelligent Algorithm (CBIA) is proposed, which seeks to improve the recognition of jaw movements using acoustic signals, even in noisy environments (e.g. inside a barn). This method is based on concepts and techniques derived from signal processing and machine learning areas to analyze the sound signal derived from ruminant feeding behavior. Two databases obtained in different experimental conditions are used to test the algorithm. The computational cost and a cost-benefit analysis of its implementation are also evaluated for its future real-time execution in a low-cost embedded system.

2. Material and methods

Two independent sound databases of ruminant feeding behaviors were used to evaluate the performance of the proposed system. One of the databases was obtained under controlled experimental conditions, showing a high SNR. This database was previously used to evaluate other algorithms, which allows a fair comparison with the algorithm introduced in this work. The second database was obtained in a barn environment, showing a poor SNR due to the presence of noises and reverberations. This database was used to evaluate the influence of the proposed methods under adverse acoustic environmental conditions. During the experiments several signal processing and machine learning techniques were evaluated.

2.1. Databases

The first database (referred as DB1) is the same used by Milone et al. (2012) and Chelotti et al. (2016) for testing CBHMM and CBRTA algorithms, respectively. DB1 was obtained at Campo Experimental J. Villarino, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario (Argentina) during February 2004. The protocols were previously evaluated and approved by the Committee on Ethical Use of Animals for Research of the Universidad Nacional de Rosario. Sound signals from dairy cows grazing either pure alfalfa (Medicago sativa) or pure fescue (Festuca arundinacea) microswards at two height levels (tall, 24.5 ± 3.8 cm, or short, 11.6 ± 1.9 cm) were recorded individually in grazing sessions conducted over a 5-day period. Forage species were selected because they differ in sward structure, water content and neutral detergent fiber content (alfalfa, 360 ± 11 g/kg and fescue, 631 ± 6 g/kg), which are factors that have a direct influence on chewing sounds (Duizer, 2001). Two 4–6 year-old lactating Holstein cows weighing 608 ± 24.9 kg, previously tamed and trained, were used. Three wireless microphones (Nady 151 VR, Nady Systems, Oakland, CA, USA) were randomly assigned to animals each day. The microphone was placed facing inwards on the forehead and was protected by a rubber foam (Milone et al., 2009). The distance between the wireless microphone and the receiver was 2-3 m. Micro-swards were established using alfalfa or fescue sown in 4-liter plastic pots, which were attached to a base-board placed inside a barn. Plants were in a vegetative state, and were intentionally manipulated to generate microswards that cows could eat with negligible displacement. The sounds were recorded at 44.1 kHz sampling frequency, 16-bit resolution and WAV format. A total of 50 grazing sessions were recorded: 15 from tall alfalfa, 11 from short alfalfa, 12 from tall fescue and 12 from short fescue. Around 50 min of acoustic signals were considered, which approximately corresponds to 3000 jaw movements (13% bites, 64% chews, and 23% chew-bites).

The signals belonging to the second database (referred as DB2) were obtained by another field experiment conducted at Campo Experimental J. Villarino, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario

(Argentina) during October of 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 behavior of five 3–5 year-old Holstein lactating cows, weighing 570 ± 40 kg, grazing alfalfa and fescue mixed pastures, were continuously monitored using a commercial recorder (Sony ICDPX312) to obtain 24 h sound recordings during six days. Sounds of biting and chewing were recorded using a directional microphone mounted over the forehead and covered by an elastic band fastened to a halter, where a recorder was attached. The signals were recorded at 44.1 kHz sampling frequency, 16-bit resolution and WAV format. Five microphone/recorder devices were randomly assigned to the cows and rotated over the six days. The signals were recorded under different environmental conditions, some of which were adverse. Noise sources included machines, noise from other animals and vehicles, among others. In addition, when the cows were inside the milking barn, sounds were affected by acoustic room reverberation.

All signals used in this study were aurally segmented and labeled by two experts in animal behavior, in order to identify and classify individual events (bite —B—, chew —C— and chew-bite —CB—) during grazing. The labeling process was done by one expert, and the result was checked by the other expert[1]. When there was disagreement, both experts worked together to provide a final decision. Because DB2 is a very large database, only those segments recorded inside the barn were selected. Twelve 5-min segments were randomly selected to test the algorithm under these adverse conditions, as hand labeling of a database of this size is unfeasible for a human expert. A typical 5-min segment contains about 300 jaw movements, which results in more than 3000 events available in 1-hour of recording. Thus, the number of required events to train and test the proposed algorithm is acceptable for the objective of testing the robustness of the algorithm on jaw movement recognition.

2.2. Chew-Bite Intelligent Algorithm

A pattern recognition system is an automatic system that aims at classifying input data into a set of specific classes using its properties and features (Duda et al., 2012). This system can be described by a series of generic

[1]Two experts decoded records of signals. Detections agreed in 100% for bites, 98.2% for chews, and 99.1% for chew-bites.

stages that allow i) the *description* of the input signal, which facilitate the extraction of distinctive features, and (ii) its *classification*, which enables identification of patterns. A block diagram of the proposed algorithm CBIA is shown in Figure 1. It shows the relationship between a typical pattern recognition system and the different stages of the algorithm: signal conditioning, preprocessing, event detection, feature extraction, and event classification.

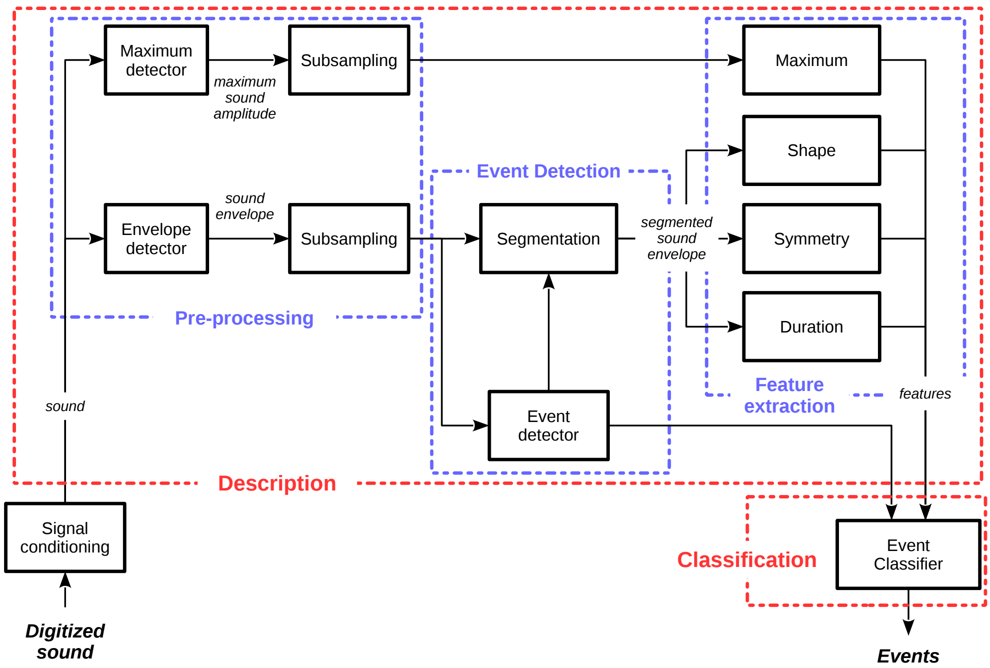


Figure 1: Complete block diagram of the proposed CBIA system.

The input of the system is the digitized sound (Figure 1) which can come from a file or an analog-to-digital converter, depending on the implementation. Within the signal conditioning stage, the numerical representation of the digitized audio is normalized and its range is matched with the range of the computer where the system is running. Sound signals sometimes show slow time-varying noises added to the target signal, especially in barn environments. Therefore, for CBIA we proposed and evaluated some detrending techniques to remove the non-stationary noises at the signal conditioning stage.

Within the pre-processing stage, the sound signal follows two paths (Figure 1):

- A maximum detector computes the maximum amplitude of the sound signal over a sliding window whose length is half of the duration of a typical chew-bite event; and
- An envelope detector that computes the sound envelope using a synchronous demodulation and a linear time-invariant (LTI) low-pass filter.

Since the sound envelope is low frequency, the signals computed by both detectors are downsampled to reduce the amount of data processed by the remaining stages. The events (potential jaw movements) are detected by comparing the sound envelope with a time-varying threshold (Chelotti et al., 2016). Then, the sound envelope is segmented and it is used to compute the features. Event features are extracted over a 1-second time-window centered where the event was located. Finally, these features are the inputs of the classifier that recognizes the event, which is the output of the system.

2.2.1. Signal conditioning

Recorded sound signals can show slow varying patterns that are added to the target signal, specially in barn environments. These trends must be removed because they can hamper the detection and classification of the events. In CBRTA no signal processing was performed over the sound signal to remove noises and trends (Chelotti et al., 2016). Hence, CBRTA left low frequency and time-varying noises unaffected when the envelope was computed. These time-varying noises modify the base level of the sound envelope, which is used to complete: i) the event detection and ii) the event classification. During the event detection, adaptive thresholds are used to detect the occurrence of a possible jaw movement by identifying peaks in the sound envelope. Regarding classification, the maximum sound amplitude could be modified by a trend. Thus, the presence of trends can modify the base level of the sound envelope, causing detection and classification problems.

In this study, three techniques were considered for the signal conditioning stage: high pass filter, least mean square filter, and empirical mode decomposition. The main objective of this stage is to attenuate the effects of trends and any other noise that can deteriorate the performance of the algorithm. The simplest signal conditioning technique is a LTI high-pass filter (HPF).

A second-order Butterworth filter was selected for the HPF and its cut-off frequency was fixed at 0.05 Hz. On the other hand, an adaptive filter has spectral characteristics controlled by time-varying parameters and a means to adjust those parameters according to an optimization algorithm. The Least Mean Squares (LMS) filter (Widrow et al., 1975) is the most popular for noise cancellation and detrending. The coefficients $W(k)$ of the filter are updated as

$$W(k+1) = W(k) + 2\mu\epsilon(k)X(k), \tag{1}$$

where $X(k)$ is the input vector at time k, μ is a factor that controls the updating rate of $W(k)$, and $\epsilon(k)$ is the error defined as the difference between the desired response $d(k)$ and the actual response $y(k)$, computed as follows

$$\epsilon(k) = d(k) - y(k) = d(k) - W^T X(k). \tag{2}$$

The parameter μ of the LMS filter was fixed at 0.01 such that a quick and stable response of the filter is obtained for all operational conditions.

Empirical Mode Decomposition (EMD) was introduced by Huang et al. (1998) for non-stationary and nonlinear signal processing, which can be used as a detrending filter. It represents signals as sums of zero-mean AM-FM components called Intrinsic Mode Functions (IMFs). An iterative algorithm called *sifting process* extracts locally the highest frequency oscillations out of original signal $x(k)$ for each mode (see Flandrin et al. (2004) for a detailed description). A fine to coarse reconstruction discarding the last modes of the decomposition (lowest frequency modes) can be performed to implement adaptive detrending filtering. Thus, the filtered signal $\widehat{x}(k)$ can be expressed as follows (Moghtaderi et al., 2013)

$$\widehat{x}(k) = \sum_{j=1}^{N_m} y_j(k), \tag{3}$$

where N_m is the number of IMFs $y_j(k)$ considered in reconstruction. It was observed that only the first seven modes contributed significantly to recognize the targeted events.

2.2.2. Feature extraction

Once the candidate events are detected, their features are extracted over a time window centered at the sample where the event was detected. Several features could be extracted (e.g. spectral, temporal or statistically derived ones), however in CBIA we propose to use a set of four temporal features that are low-cost and with discriminative power for this problem:

- **Shape index**: is computed as the number of zero-crossings in the sign of the derivative signal obtained from the envelope signal (third row in Figure 2). This calculation is performed only if envelope amplitude exceeds a noise threshold. This feature provides useful information to differentiate simple events (chew and bite) from combined events (chew-bite).

- **Maximum intensity**: provides information to differentiate low-amplitude events (chews) from high-amplitude ones (bites and chew-bites). This feature is computed directly from the sound signal over a sliding window with length equal to the period of a chew-bite event (fourth row in Figure 2).

- **Duration**: is computed as the time period in which amplitude of the envelope is greater than a given threshold. In general, the duration of compound events (chew-bite) is larger than simple events (chew or bite), which are similar (fifth row in Figure 2).

- **Symmetry**: is computed as the ratio between the left area and the total area of the event. Left and right event areas are divided at the first peak of the event (last row in Figure 2). It can provide discriminative information because events have different symmetries.

2.2.3. Classification

Classification of ingestive events using sounds is a well-defined problem. For this stage, six classifiers were considered: i) decision tree, ii) random forest, iii) multilayer perceptron, iv) radial basis function network, v) support vector machine, and vi) extreme learning machine. The most relevant parameters of classifiers were optimized using the grid search method.

Decision Trees (DTs) have the ability of learning simple decision rules and systematizing them in order to arrive at complex decisions. They were built using the C4.5 algorithm and pruning confidence was optimized (Breiman et al., 1984; Ross Quinlan, 2014).

Multilayer Perceptron (MLP) is a conventional feed-forward artificial neural network design that can deal with non-linearly separable data (Bishop,

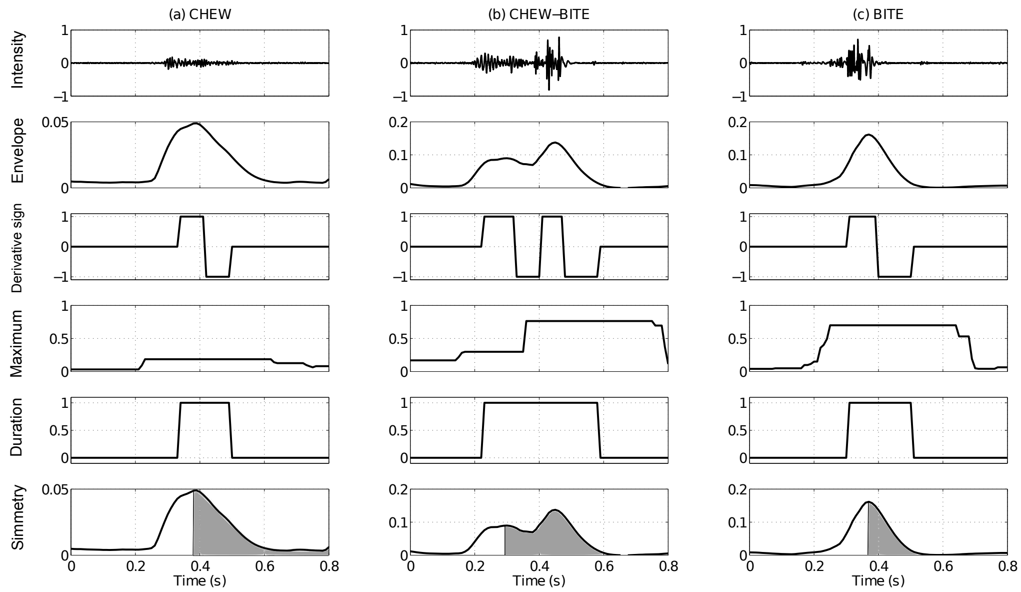


Figure 2: Typical acoustic events produced by jaw movements and derived signals from where the features are extracted. From top to bottom: i) acoustic signal, ii) sound envelope, iii) shape index, iv) maximum intensity, v) duration and vi) symmetry.

2006). They were configured with four inputs neurons (number of input features) and three outputs neurons (number of output labels). Features were normalized and output labels binarized to match MLP output. MLPs with one and two hidden layers were tested. The number of neurons in each hidden layer and the learning rate were optimized. Typically, four hidden neurons were chosen.

Radial Basis Function (RBF) network is another type of artificial neural network (Bishop, 2006). The number of RBF hidden-layer units was optimized. Random Forest (RF) is an ensemble of decision trees which generally results in an overall better model. The number of trees was optimized during training.

For the support vector machine (SVM), a radial basis function was chosen as a kernel and a soft margin penalty for misclassifications was considered (Steinwart and Christmann, 2008; Hastie et al., 2009). Penalty coefficient C and parameter γ of SVMs were optimized. The one-against-one approach was followed for the multiclass classification task with SVMs (Hall et al., 2009; Chang and Lin, 2011).

Extreme learning machine (ELM) is a new type of neural network (Huang et al., 2006), which has shown good generalization performance and short training times. The number of hidden neurons and regularization parameter γ of ELM were optimized (Deng et al., 2009). Finally, the optimization was performed using the training set of each fold in order to preserve cross-validation (Hall et al., 2009).

2.3. Experimental setup

In this study, leave-one-signal-out cross-validation was used to conduct the experiments. In the first fold of this scheme, one signal was taken for testing, while the remaining signals were used for selecting the best parameters of the classifier and for training the models. The following folds switch the test signal to another one until all signals are considered in the test. Reported performance metrics were obtained averaging across folds. They are recognition rate (RR), recall (R), and precision (P),

$$\begin{aligned} RR &= \frac{\sum_i tp_i + tn_i}{\sum_i tp_i + tn_i + fp_i + fn_i}, \\ R &= \frac{\sum_i tp_i}{\sum_i tp_i + fn_i}, \\ P &= \frac{\sum_i tp_i}{\sum_i tp_i + fp_i}, \end{aligned} \tag{4}$$

where tp_i are true positives, tn_i true negatives, fp_i false positives, and fn_i false negatives counts for class i, respectively (Sokolova and Lapalme, 2009). Models were pasture-specific, thus classification was performed with the corresponding specific model and reported accordingly, like CBHMM (Milone et al., 2012). Required computational load was also compared.

3. Results

The following subsections present a detailed analysis of results obtained. First, a graphical analysis of proposed features is presented. Second, variants of the proposed algorithm are analyzed based on their discriminative power. Finally, the proposed algorithm is compared with state-of-the-art techniques based on their performance metrics and computational costs.

3.1. Analysis of proposed features

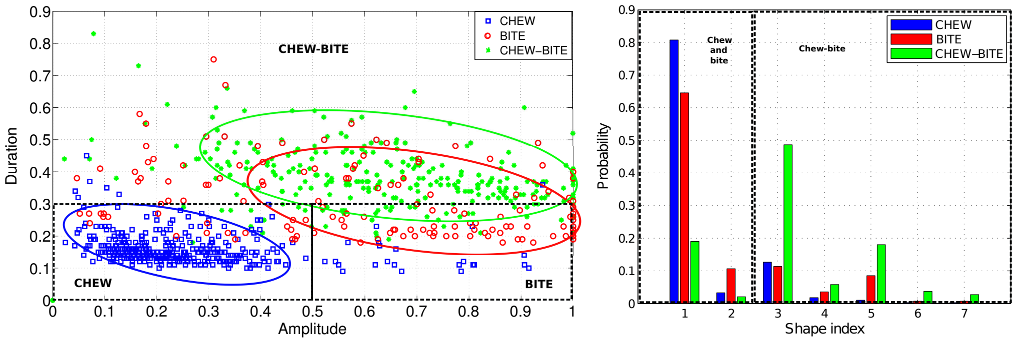
Figure 3 shows the data and the decision regions generated by the rules employed in the CBRTA (Chelotti et al., 2016). The data of chews and chew-bites is organized in dense clusters (highlighted by ellipses), while the data of bites is scattered in the plane (Figure 3.a). The majority of chews reach a shape index of one or two, while bites and chew-bites can have indices greater than one (Figure 3.b). A simple analysis shows that chews can be clearly classified, while the classification of bites and chew-bites is more difficult due to the lack of discriminative information (Figure 3.a). This analysis is consistent with the results from Chelotti et al. (2016), where chews are classified with an accuracy over 90% while chew-bites and bites are classified with an accuracy of 67% and 84%, respectively. The need for the introduction of an additional feature that allows to clearly separate bites from chew-bites is evident.

Figure 4 shows the data distribution of amplitude and duration against symmetry (a new feature not previously used in CBRTA). In this case, the data exhibits a highly clustered organization with minimal overlapping between classes. A symmetry around 0.5 will indicate a symmetric event in terms of this feature. In the case of chew-bites this relation is lower than 0.5 since the area is divided at the first peak (last row in Figure 2), while for bites this feature takes values greater than 0.5 (Figure 4). This new discriminative information, combined with shape index and duration, should allow to differentiate between bites and chew-bites.

3.2. CBIA recognition results

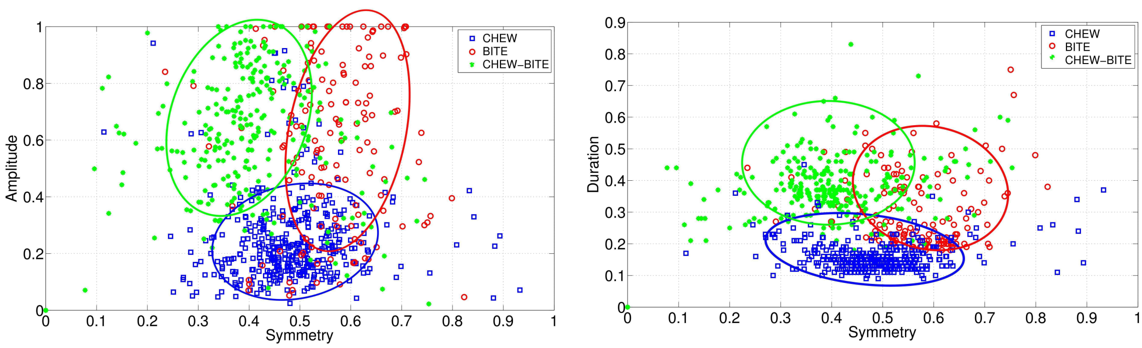
The proposed system can combine different detrending and classification techniques. A summary of average recognition rates for all combinations is presented in Figure 5. The rates were obtained averaging across all signals in DB1. In addition to the evaluated detrending techniques (HPF, LMS, and EMD), features were extracted from raw signals (RS), i.e. signals to which no detrending method was applied. Most combinations exceeded 80% recognition rate. ELM combinations achieved the poorest results compared to other classifiers. Application of a detrending stage resulted in clear improvements over the raw signals, where adaptive variants (LMS or EMD) reached recognition rates above 90%.

Average recognition rates for each pasture are detailed in Table 1. In general, recognition rates were above 80% except for short alfalfa, which was the most difficult pasture for event recognition. Between the two pastures



(a) (b)

Figure 3: Data distribution for tall alfalfa (DB1) and decision regions of CBRTA: (a) amplitude vs duration and (b) shape index.



(a) (b)

Figure 4: Data distribution for tall alfalfa (DB1): (a) amplitude and (b) duration against new CBIA symmetry feature.

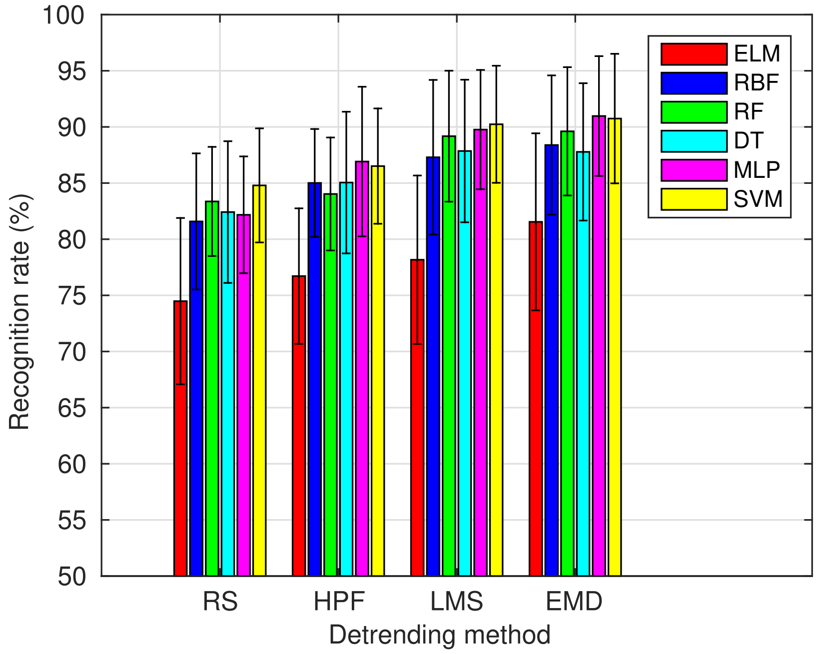


Figure 5: Average recognition rates (%) of CBIA across all signals in DB1 for different combinations of detrending and classification techniques.

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
J. O. Chelotti, S. R. Vanrell, J. Galli, L. Giovanini & H. L. Rufiner; "A pattern recognition approach for detecting and classifying jaw movements in grazing cattl"
Computers and Electronics in Agriculture, Vol. 145, pp. 83-191, 2018.

evaluated (alfalfa and fescue) the best results were obtained for fescue, in some cases exceeding a 90% recognition rate. Among detrending techniques, LMS and EMD produced the largest improvements, whereas HPF showed small improvements. Among the evaluated classifiers, MLP, SVM and RF achieved the best results, while ELM generated the poorest results.

In order to confirm the advantages of a detrending stage, a more challenging scenario was developed using signals from DB2. Cattle barn is a known noisy environment. Hence, a random selection of 5-min segments from grazing signals recorded in this environment was performed. Because these signals were continuously recorded from mixed pastures in different experimental conditions, the classifiers were retrained. The average recognition rates obtained on these segments are summarized in Table 2. Only the techniques that achieved the best results on DB1 were considered for experiments on DB2. In this sense, the adaptive methods (LMS and EMD) were evaluated for the signal conditioning stage and three classifiers (MLP, SVM and DT) were considered for the classification stage, based on recognition

Table 1: Average recognition rates (%) of CBIA obtained for each pasture of DB1 for different combinations of detrending and classification techniques. Bold numbers indicate the best results.

	Pasture	MLP	SVM	DT	RF	ELM	RBF
RS	Tall alfalfa	84.58	85.04	84.74	**85.57**	81.81	83.60
	Short alfalfa	76.55	77.27	76.20	**77.92**	72.86	73.88
	Tall fescue	88.34	**88.47**	87.47	86.70	78.99	86.52
	Short fescue	86.42	**88.00**	84.12	86.22	80.40	85.19
HPF	Tall alfalfa	84.39	86.14	85.42	**86.46**	82.18	85.72
	Short alfalfa	76.47	78.50	75.93	**78.54**	73.46	75.00
	Tall fescue	89.34	**89.71**	88.55	87.03	76.67	87.74
	Short fescue	88.98	**90.12**	85.04	86.93	80.95	86.11
LMS	Tall alfalfa	87.13	87.06	86.11	**87.41**	83.63	85.46
	Short alfalfa	79.13	**80.83**	78.50	80.46	76.29	78.17
	Tall fescue	89.85	**90.18**	87.97	89.30	80.28	86.52
	Short fescue	**90.14**	90.05	85.79	86.54	83.06	86.05
EMD	Tall alfalfa	**88.01**	87.52	85.00	85.97	84.12	86.76
	Short alfalfa	79.71	**82.00**	79.09	81.23	75.97	78.36
	Tall fescue	**90.29**	89.99	87.84	89.46	76.83	88.46
	Short fescue	**93.17**	92.04	86.61	89.59	84.45	86.81

rates presented in Table 1 and the simplicity of their implementation.

Table 2: Average recognition rates (%) of CBIA for adaptive detrending methods and selected classifiers on 5-min segments from DB2. Bold numbers indicate the best results.

	MLP	SVM	DT
RS	74.79	74.52	73.56
LMS	80.72	**81.61**	79.21
EMD	**82.27**	80.91	**80.23**

Baseline results in Table 2 were obtained on raw signals (RS) and the use of adaptive detrending showed an improvement of 5.65% or greater. The combination EMD+MLP achieved the best result (82.27%) on these signals, surpassing MLP baseline results by 7.48%. The second best combination was LMS+SVM (81.61%), with an improvement of 7.09% over RS for SVM classifier. The combinations LMS+DT and EMD+DT also achieved very good recognition rates.

Table 3: Comparison of performance measures (%) for different algorithms (including state-of-the-art methods) applied on DB1 and averaged over all pastures. Bold numbers indicate the best results.

	CBHMM	CBRTA	CBIA			
			LMS+DT	LMS+MLP	LMS+SVM	EMD+SVM
Recognition rate	79.50	77.00	87.85	89.76	90.23	**90.74**
Recall	83.27	70.76	88.16	91.63	90.88	**92.57**
Precision	85.09	83.85	86.35	89.75	89.51	**92.21**

Table 4: Comparison of computational costs for different algorithms given in terms of operations per second of recorded audio signal.

Algorithm stage	CBHMM†	CBRTA	CBIA			
			LMS+DT	LMS+MLP	LMS+SVM	EMD+SVM
Signal conditioning and pre-processing	NA	NA	10,000	10,000	10,000	8,991,943
Detection and feature extraction	2,122,857	27,700	28,800	28,800	28,800	28,800
Classification	53,557,392	30	20	160	1600	1600
Global cost	55,680,249	27,730	38,820	38,960	40,400	9,022,343

† CBHMM does not perform event detection, features are extracted continuously.

3.3. Comparison with existing methods

A comparison between proposed (CBIA) and former methods (CBHMM and CBRTA) is presented in Table 3. The results obtained on DB1 are averaged over all pastures and summarized by recognition rate, recall and precision. Adaptive detrending methods and classifiers of CBIA greatly improved the recognition rate over former algorithms. Recall and precision were above 85% for all variants of the proposed method.

Table 4 shows a detailed analysis of the computational costs for each stage of the algorithms (former and proposed). The assumptions to estimate these costs are detailed in Appendix A. With the exception of CBHMM, the classification stage has a relatively low computational cost compared to preceding stages. Signal detrending performed with EMD requires 900 times more computations than LMS. Regarding classification, SVM cost is several times higher than heuristic rules, MLP, or DT costs. CBIA global cost using LMS has the same order of magnitude as CBRTA. By contrast, CBHMM and CBIA with EMD are several orders of magnitude more costly than other CBIA variants.

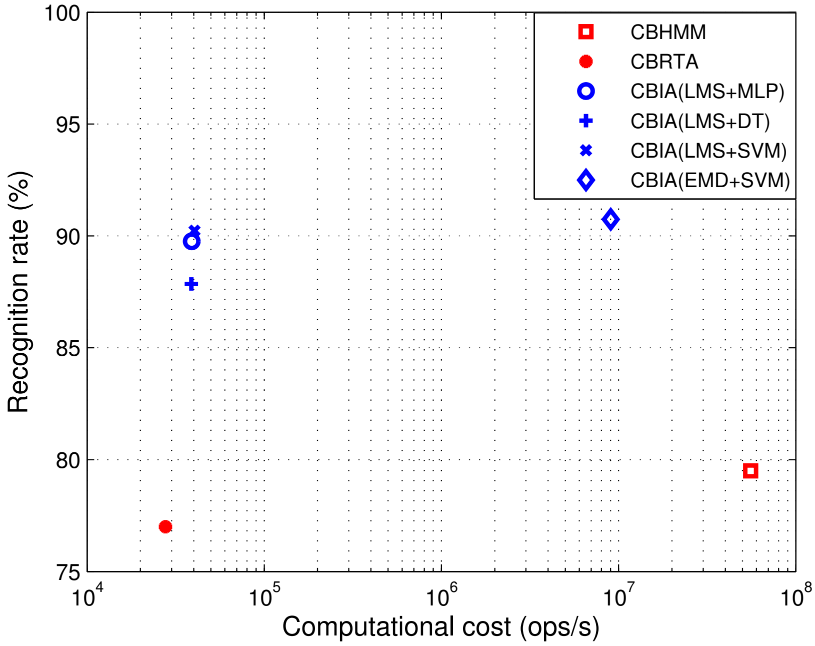


Figure 6: Recognition rate vs computational cost for different algorithms.

Figure 6 illustrate the trade-off between recognition rate and computational cost for the methods presented in Tables 3 and 4. CBHMM is the most computationally expensive algorithm and achieved almost 80% recognition rate. By contrast, CBRTA has the lowest computational cost and recognition rate. Among the proposed methods, the one that used EMD produced very good recognition rates, but its computational cost is very high. CBIA variants with LMS achieved very good recognition rates and have relatively low computational cost.

4. Discussion

CBIA achieved good recognition rates for several variants of the algorithm. Better performance rates were obtained for DB1 than for DB2, due to the higher SNR of DB1 signals. However, the benefit of a detrending stage was more evident on DB2 (Table 2). A comparison of the CBIA variants with previous algorithms shows that better performance rates are achieved with a minor increment in computational cost (Figure 6).

Regarding the detrending stage of the proposed algorithm, the adaptive techniques (EMD and LMS) outperformed the fixed high-pass filter (HPF), probably due to the fact that many of the noise sources seem to be non-stationary. In this sense, EMD achieved the best results, reaching up to 93% recognition rate, but its computational cost was extremely high. Related algorithms such as ensemble empirical mode decomposition (Wu and Huang, 2009) were not taken into account in this work because they usually have even higher computational cost. In contrast, LMS showed a performance almost as good as EMD for several of the combinations evaluated (LMS+MLP, LMS+SVM), with a lower computational cost. This establishes LMS as the most appropriate technique for real-time execution of the algorithm.

Six classifiers were evaluated for the classification stage. From the standpoint of recognition, the best results correspond to SVM, RF and MLP, with a recognition rate of about 90% in some cases. From the computational cost point of view, the best classifiers were DT, MLP and SVM. While SVM generally achieves the best classification results, DT and MLP are easily implemented in an embedded system. DT has the additional advantage that its rules allows an easy interpretion and understanding on how the classification process is performed. MLP can define arbitrary decision boundaries (e.g. non-linear boundaries) and thus obtain higher recognition rates at the expense of a computational cost slightly higher. For this reason, the cost-benefit analysis presents LMS+MLP as the best combination for CBIA implementation.

Some interesting similarities were observed in the comparison between CBRTA heuristically derived rules (RR: ∼77%) and CBIA-DT automatically learned rules (RR: ∼88%). Both strategies share a low computational cost and easy interpretation. CBRTA rules are simpler and require a small set of comparisons for classifying an event. In contrast, CBIA-DT rules require two to four times more comparisons in order to perform the same classification. Regarding rule definition, in a multiclass problem, the selection of a CBIA-DT split takes all classes into account. By contrast, the strategy to obtain heuristic rules concentrates on one class at a time, disregarding what happens to the other classes. CBRTA uses its rules separately for each type of event, whereas CBIA-DT share some rules among classes. Regarding selected features and their split values, CBRTA divided the feature space into rough regions, while CBIA-DT performed a sharper division using similar split values in combination with new ones. For instance, event duration was selected as the most important feature and its split values were around 0.3 s, like it was defined for the CBRTA (see Figure 3.a).

Probably, changes in forage characteristics such as water content, anatomy of tissues, and fiber content, and animal characteristics such as dentition, head size and anatomy will tend to affect the sound produced by the ingestion of forage (Galli et al., 2017). In this sense, short alfalfa was the pasture with the worst results in recognition (see Table 1), which is consistent with the results obtained by Milone et al. (2012) and Chelotti et al. (2016). A possible explanation for this is that short alfalfa plants have a higher ratio of stems to leaves than tall alfalfa and fescue. This could produce bite sounds with lower amplitude, increasing confusion between events and thus affecting the classification stage. Some of these problems could be solved by designing specific electronic circuits of acquisition and analog signal conditioning as part of the system. However, the commercial recorders used in this work do not allow such attempt. A first approach considering this idea has been presented in Deniz et al. (2017). On the other hand, the results in the present study show that for animals of similar size, age and breed, CBIA was able to detect and classify jaw movements in grazing cattle with high accuracy.

5. Conclusions

In this study an algorithm for detection and classification of masticatory events from acoustic signals was presented. The proposed CBIA consists of five stages: i) signal conditioning, ii) pre-processing, iii) event detection, iv) feature extraction, and v) event classification. Within the conditioning stage, three detrending methods were evaluated. Adaptive thresholding was used for event detection and an improved set of discriminative features was extracted. Classical and advanced machine learning techniques were used for classification of three types of masticatory events (chew, bite and chew-bite). The best trade-off between recognition rate and computational cost was obtained with LMS+MLP variant. The high recognition rates achieved (up to 90% and over 80%) for databases recorded in different experimental conditions, demonstrate the robustness of this approach. CBIA outperformed previous methods (CBHMM and CBRTA) by at least 10% in terms of the recognition rate. Many of the evaluated variants have the additional advantage of a low computational cost, which allows real-time execution.

In this study, models were trained and evaluated in a pasture-specific fashion. Most of the time the type of pasture can be known in advance, thus this is not a real limitation for the system. Future research will focus on recognizing masticatory events without this assumption, in a pasture independent way, which may require different system architecture or strategies.

Acknowledgements

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, project PICT 2011-2440, 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).

References

Alkon, P., Cohen, A., 1986. Acoustical biotelemetry for wildlife research: a preliminary test and prospects. Wildl. Soc. Bull. 14, 193–196.

Alkon, P. U., Cohen, Y., Jordan, P. A., 1989. Towards an acoustic biotelemetry system for animal behavior studies. J. Wildl. Manage. 53 (3), 658.

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. Biotechnology, Agronomy, Society and Environment 20.

Bishop, C. M., 17 Aug. 2006. Pattern Recognition and Machine Learning. Springer.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A., 1984. Classification and Regression Trees. Chapman and Hall/CRC.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM. ACM Transactions on Intelligent Systems and Technology 2 (3), 1–27.

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.

De Boever, J. L., Andries, J. I., De Brabander, D. L., Cottyn, B. G., Buysse, F. X., 1990. Chewing activity of ruminants as a measure of physical structure — a review of factors affecting it. Anim. Feed Sci. Technol. 27 (4), 281–291.

Deng, W., Zheng, Q., Chen, L., 2009. Regularized extreme learning machine. In: IEEE Symposium on Computational Intelligence and Data Mining.

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.

Duda, R. O., Hart, P. E., Stork, D. G., 9 Nov. 2012. Pattern Classification. John Wiley & Sons.

Duizer, L., 2001. A review of acoustic research for studying the sensory perception of crisp, crunchy and crackly textures. Trends in food science & technology 12 (1), 17–24.

Flandrin, P., Rilling, G., Goncalves, P., 2004. Empirical mode decomposition as a filter bank. IEEE Signal Processing Letters 11 (2), 112–114.

Galli, J., Cangiano, C., Pece, M., Larripa, M., Milone, D., Utsumi, S., Laca, E., 2017. Monitoring and assessment of ingestive chewing sounds for prediction of herbage intake rate in grazing cattle. animal, 1–10.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., 2009. The WEKA data mining software. ACM SIGKDD Explorations Newsletter 11 (1), 10.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Science & Business Media.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70 (1-3), 489–501.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 454 (1971), 903–995.

Klein, L., Baker, S., Purser, D., Zaknich, A., Bray, A., 1994. Telemetry to monitor sounds of chews during eating and rumination by grazing sheep. In: Proceedings-australian Society Of Animal Production. Vol. 20. Australian society of animal production, pp. 423–423.

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 Forage Sci. 47 (1), 81–90.

Milone, D., Rufiner, H., Galli, J., Laca, E., Cangiano, C., 2009. Computational method for segmentation and classification of ingestive sounds in sheep. Computers and Electronics in Agriculture 65 (2), 228–237.

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.

Moghtaderi, A., Flandrin, P., Borgnat, P., 2013. Trend filtering via empirical mode decompositions. Comput. Stat. Data Anal. 58, 114–126.

Nadin, L. B., Chopa, F. S., Gibb, M. J., Trindade, J. K. d., Amaral, G. A. d., Carvalho, P. C. d. F., Gonda, H. L., 2012. Comparison of methods to quantify the number of bites in calves grazing winter oats with different sward heights. Applied Animal Behaviour Science 139 (1), 50 – 57.

Navon, S., Mizrach, A., Hetzroni, A., Ungar, E. D., 2013. Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring. Biosystems Eng. 114 (4), 474–483.

Nelson, D. E., Alkon, P. U., Krausman, P. R., 2005. Using acoustic telemetry to monitor foraging by penned mule deer. Wildlife Society Bulletin 33 (2), 624–632.

Oudshoorn, F. W., Cornou, C., Hellwing, A. L. F., Hansen, H. H., Munksgaard, L., Lund, P., Kristensen, T., Nov. 2013. Estimation of grass intake on pasture for dairy cows using tightly and loosely mounted di- and triaxial accelerometers combined with bite count. Computers and Electronics in Agriculture 99, 227–235.

Ross Quinlan, J., 28 Jun. 2014. C4.5: Programs for Machine Learning. Elsevier.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information Processing & Management 45 (4), 427–437.

Steinwart, I., Christmann, A., 15 Sep. 2008. Support Vector Machines. Springer Science & Business Media.

Tani, Y., Yokota, Y., Yayota, M., Ohtani, S., Mar. 2013. Automatic recognition and classification of cattle chewing activity by an acoustic monitoring method with a single-axis acceleration sensor. Computers and Electronics in Agriculture 92, 54–65.

Ungar, E. D., Ravid, N., Zada, T., Ben-Moshe, E., Yonatan, R., Baram, H., Genizi, A., 2006. The implications of compound chew–bite jaw movements for bite rate in grazing cattle. Applied animal behaviour science 98 (3), 183–195.

Wang, Y.-H., Yeh, C.-H., Young, H.-W. V., Hu, K., Lo, M.-T., Apr. 2014. On the computational complexity of the empirical mode decomposition algorithm. Physica A: Statistical Mechanics and its Applications 400, 159–167.

Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, J. E., Goodlin, R. C., 1975. Adaptive noise cancelling: Principles and applications. Proc. IEEE 63 (12), 1692–1716.

Wu, Z., Huang, N. E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in adaptive data analysis 1 (01), 1–41.

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.

Appendix A. Computational costs

Computational costs of different algorithms were presented in Table 4 for each stage. Practical and typical values were considered in order to get a straightforward comparison. Computations required are estimated for processing one second of audio signal, which sampling frequency was fixed at 2 kHz ($n = 2000$ samples). Event detection, feature extraction and classification for CBHMM and CBRTA were analyzed in Chelotti et al. (2016). The CBHMM requires:

$$24(21 + 26 \cdot 1.5n + 1.5n \log(1.5n)) + 53,557,392 = 55,680,249 \text{ ops.} \quad \text{(A.1)}$$

It is the most computationally expensive algorithm of those analyzed in this study. In fact, CBHMM does not perform event detection, features are extracted with a sliding window, and then classification is accomplished.

By contrast, CBRTA and CBIA perform event detection prior to feature extraction, and then the corresponding classification of detected events. The event detection typically yields 1 or 2 events for each second of signal[2], and in the following it is assumed the worst case, 2 events per second. Detection and feature extraction cost of CBRTA is:

$$13n + 1,700 = 27,700 \text{ ops.} \tag{A.2}$$

To this cost it is necessary to add the cost of the evaluation of classification rules for the three events (30 operations). CBIA has a similar cost for detection and feature extraction (27,800 operations) compared to CBRTA, but it requires extra symmetry computation (100 operations).

The cost of CBIA detrending techniques was estimated for LMS and EMD. LMS requires 5 operations per signal sample, thus, it requires 10,000 operations to process one second of audio signal, while EMD requires at least:

$$41 N_S \; n \; \log_2(n) \tag{A.3}$$

operations for extracting all IMFs (Wang et al., 2014), where the typical number of siftings $N_S = 10$ was chosen, which gives a total of 8,991,943 operations per second of signal.

Classification cost of CBIA was evaluated for DT, MLP, and SVM. The number of features (4) and the number of classes (3) were considered. DT is the simplest one requiring 10 operations per event (typical depth of 10). MLP has a cost of 80 operations per event, which is directly related to the number of input nodes (4), the number of hidden layers (1), the number of hidden neurons (4 typically), and the number of outputs (3). Regarding SVM, an average of 200 support vectors were selected to build classifiers. Then, as prediction complexity of SVM is proportional to the number of support vectors and the number of features, it requires 800 operations per event, at least.

[2]This is derived from the frequency of typical events of cattle feeding behavior.