An online method for estimating grazing and rumination bouts using acoustic signals in grazing cattle

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

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

Abstract

The growth of the world population expected for the next decade will increase the demand for products derived from cattle (i.e., milk and meat). In this sense, precision livestock farming proposes to optimize livestock production using information and communication technologies for monitoring animals. Although there are several methodologies for monitoring foraging behavior, the acoustic method has shown to be successful in previous studies. However, there is no online acoustic method for the recognition of rumination and grazing bouts that can be implemented in a low-cost device. In this study, an online algorithm called bottom-up foraging activity recognizer (BUFAR) is proposed. The method is based on the recognition of jaw movements

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

from sound, which are then analyzed by groups to recognize rumination and grazing bouts. Two variants of the activity recognizer were explored, which were based on a multilayer perceptron (BUFAR-MLP) and a decision tree (BUFAR-DT). These variants were evaluated and compared under the same conditions with a known method for offline analysis. Compared to the former method, the proposed method showed superior results in the estimation of grazing and rumination bouts. The MLP-variant showed the best results, reaching F1-scores higher than 0.75 for both activities. In addition, the MLP-variant outperformed a commercial rumination time estimation system. A great advantage of BUFAR is the low computational cost, which is about 50 times lower than that corresponding to the former method. The good performance and low computational cost makes BUFAR a highly feasible method for real-time execution in a low-cost embedded monitoring system. The advantages provided by this system will allow the development of a portable device for online monitoring of the foraging behavior of ruminants. Web demo available at: https://sinc.unl.edu.ar/web-demo/bufar/

Keywords: Acoustic monitoring, activity recognition, ruminant foraging behavior, precision livestock farming, pattern recognition, machine learning.

1. Introduction

Accurate monitoring of animal foraging behavior is a complex but essential task to optimize livestock production systems (Hodgson and Illius, 1998). Changes in the ruminant foraging behavior are indicators of animal health and welfare and can be useful in early detection and prevention of several diseases. For example, an increment in rumination time can be associated with an increment of saliva production and improvements in rumen health (Beauchemin, 1991). Conversely, a reduction of rumination can be interpreted as an indicator of stress (Herskin et al., 2004), anxiety (Bristow and Holmes, 2007), or a disease (Hansen et al., 2003; Paudyal et al., 2018; Welch, 1982). In the last decade, precision livestock farming has been presented as a useful approach to tackle these problems, using advanced technology to monitor each animal. In this sense, recent technological developments have facilitated the use of sensors to monitor many physical variables both for animal science research and for practical farm level applications (Berckmans, 2014).

Foraging behavior of ruminants can be characterized by jaw movements (short timescale) and activities (long timescale). Jaw movements (JM) have a duration close to 1 s, whereas activity bouts can last from minutes to hours. The JM (or masticatory events) are biting, when herbage is apprehended and severed; chewing, when herbage is comminuted; and a combination of chewing and biting in a single JM, which is called chew-bite (Galli et al., 2018; Laca et al., 1992; Ungar and Rutter, 2006). Main foraging activities are grazing and rumination. Their duration widely fluctuates in the day. Grazing can cover from 25 to 50% of the day and rumination from 15 to 40% (Hodgson, 1990; Kilgour, 2012; Phillips, 1993). The grazing process involves searching, apprehending, chewing, and swallowing herbage. Rumination involves bolus regurgitation, chewing, and deglutition, in a periodic cycle that typically last 1 min. During both activities, JM are performed rhythmically with a frequency that ranges from 0.75 to 1.20 JM per second (Andriamandroso et al., 2016). While grazing, the three types of JM are present (i.e., chew, bite and chew-bite), whereas only chews are present during rumination (Hodgson and Illius, 1998).

An automatic monitoring system should be reliable, insightful, and practical to implement. For instance, these goals imply that recorded signals should be analyzed without human assistance, that the methodology should be scalable to large herds (even in pasture-based production systems), that the device autonomy should facilitate the collection of data over long periods of time (from days to weeks), and that data should be processed online to reduce in-device data-storing and communication requirements. Thus, an ideal methodology to be deployed in the field is one that is powerful at characterizing the foraging behavior as well as it is efficient at data processing.

Different sensing technologies have been used in the development of automatic monitoring systems, such as motion sensors, noseband pressure sensors, and microphones (Andriamandroso et al., 2016). Among motion sensors it is widespread the use of accelerometers (Arcidiacono et al., 2017; Giovanetti et al., 2017; González et al., 2015; Martiskainen et al., 2009) and inertial measurement units (Andriamandroso et al., 2017; Greenwood et al., 2017; Smith et al., 2016). These sensors have been used to recognize a broader set of activities such as rumination, grazing, resting, drinking and walking. An activity is determined by a postural analysis of the animal, where the sensors are used to estimate the position and motion of its head and body. However, this strategy can confuse activities that share the same posture. A better strategy for recognizing ruminating, eating and drinking activities is the use of noseband pressure sensors (Nydegger et al., 2010; Rutter, 2000; Rutter et al., 1997; Werner et al., 2018; Zehner et al., 2017). They have been used in the analysis of housed and free-grazing cows during one- to two-hour sessions. This yielded very good results, but further studies are required for continuous long-term monitoring. A limitation of this approach is that does not discriminate between JM (i.e., they are not classified) which is a requirement for a more detailed analysis such as herbage intake estimation (Galli et al., 2018).

Acoustic monitoring has proven to be reliable for recognizing short-term JM in free-ranging cows (Chelotti et al., 2018; Clapham et al., 2011; Laca et al., 1992; Milone et al., 2012; Navon et al., 2013). In particular, the chew-bite intelligent algorithm (CBIA) performs an online processing of the sound signal and has achieved very good results (Chelotti et al., 2018). A related commercial monitoring system is the Hi-Tag system (SCR Engineers Ltd., Netanya, Israel). Its design is focused on the autonomy, portability and hardware robustness required by the application. Besides it is based on microphones, the analysis of the signal is exclusively focused on rumination monitoring (Goldhawk et al., 2013; Schirmann et al., 2009). Recently, acoustic monitoring has also been successful on long-term recognition of foraging activities in free-ranging cows (Vanrell et al., 2018). The regularity-based acoustic foraging activity recognizer (RAFAR) was able to identify grazing and rumination bouts from sound recordings. The success of RAFAR relies on an offline analysis of long recordings (several hours), which clearly expose the regularities of foraging activities. Those recordings are acquired in each animal of the herd and then analyzed in a desktop computer. However, there are some practical limitations with this approach. A portable device, has limited storage capacity, processing capability, and power supply. These limitations becomes more relevant when the application on large herds is desired.

In this study, the acoustic monitoring strategy is taken one step further. The main point to explore is the potential of identifying the foraging activities from a prior recognition of JM following a bottom-up approach. The proposed method is focused on an online processing of the acoustic signals , i.e. the input signal is processed sample-by-sample, as it is received. In addition, the method should have relatively low computational cost and be focused on its real-time implementation in a low-cost embedded system. This would contribute to establish the acoustic monitoring as a non-invasive alternative that could handle the requirements of the application and can provide insights about natural foraging behavior of ruminants.

2. Material and methods

2.1. Proposed method

An online method for detection and classification of the most important foraging activities of ruminants is presented in this section. The method can process the signal sample-by-sample (online fashion). The bottom-up foraging activity recognizer (BUFAR) has two levels of recognition. First, JM are recognized and then this information is used to estimate rumination and grazing bouts. As a result, the information about nutritional status can be enhanced by providing statistics of both JM and activity bouts.

Fig. 1 shows typical sound recordings during (a) grazing and (b) rumination. The amplitude of the sound signals might be seen as an obvious measure for discrimination. However, variations in the amplitude across mi-

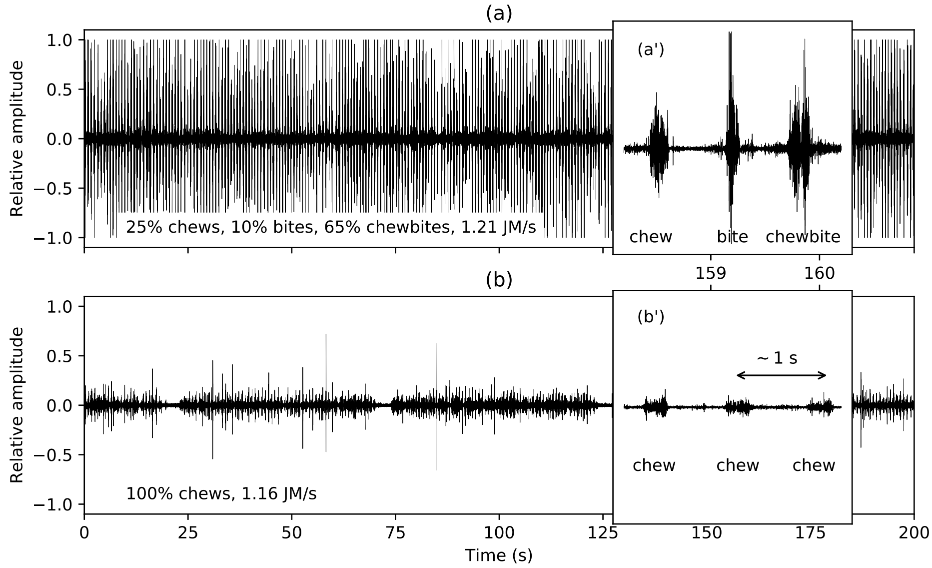

Figure 1: (a) Grazing and (b) rumination activities. Typical percentages and rate of jaw movements by activity. The jaw movement included in each activity are zoom-in.

crophones, recording devices, sessions, and cows have not allowed a reliable classification. By contrast, the rate of JM of both activities is very similar and it helps to distinguish activity bouts from noisy segments in the recordings. A clear difference between the activities is the proportion of JM. For example, in these recordings, grazing has 25% of chews, 10% of bites, and 65% of chew-bites, whereas rumination has a 100% of chews. Thus, the rate and the proportion of JM are the keys of the proposed method.

A diagram of the proposed system BUFAR is shown in Fig. 2. It has five stages that perform the required processing of data to recognize JM and foraging activities. For the sake of a low computational cost, tasks within each stage have been simplified whenever it was possible. The input of the system is the sound signal produced during foraging activities. Three activities are considered: rumination, grazing, and other activities. Other activities include any activity other than rumination or grazing (i.e., milking, silence , confusing sounds, etc.). Detection and classification of JM are performed with the CBIA algorithm (Chelotti et al., 2018). CBIA comprises three stages: signal pre-processing, jaw-movement detection, and jaw-movement classification. In signal pre-processing stage, the raw signal is conditioned and filtered to improve the signal-to-noise ratio (SNR) and remove slow varying trends. Jaw-movement detection stage spots these movements by analyzing the filtered signal with an adaptive threshold. Each JM is assigned with a timestamp and a set of features (duration, maximum amplitude, shape index, and symmetry). The timestamp is saved in the segment buffer and it will be used for activity recognition. In the classification stage, the features of each JM are taken by a neural network model to assign an event label: bite (b), chew (c), or chew-bite (cb).

The proposed system performs activity recognition by analyzing fixed-length segments of the acoustic signal. JM that are detected and classified within a segment are stored in a segment buffer. The rate of JM in a segment and the proportions of their types are computed to feed the last processing stage. At this point, activity classification could be seen as a simple task, but an exploratory data analysis on the training set has shown a complex underlying distribution of the segment features (rate, %c, %b, %cb). The rate of recognized JM during rumination and grazing is expected to be in the range from 0.75 to 1.40 Hz (Fig. 3). By contrast, the rate of JM identified during other activities presents a lower frequency. The overlapping among

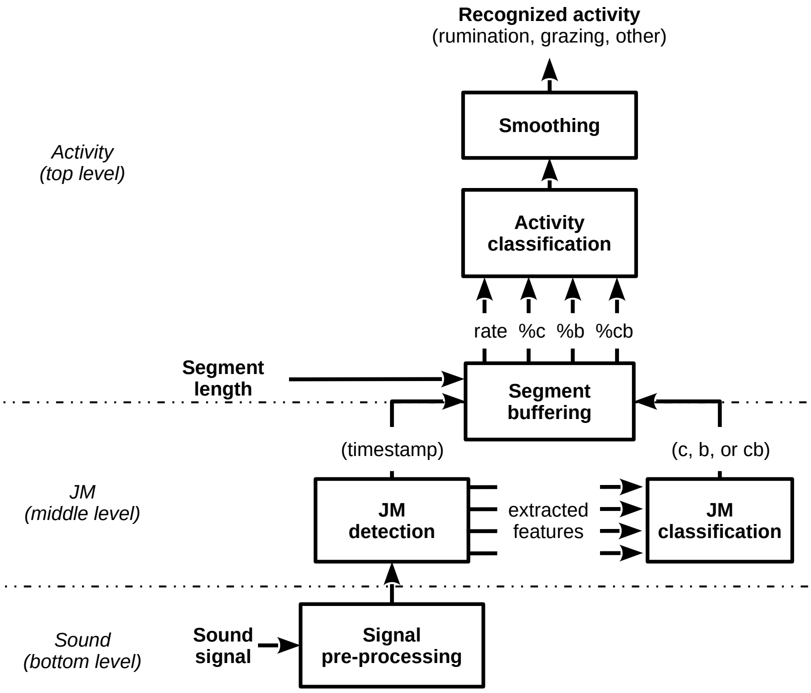

Figure 2: General diagram of the bottom-up foraging activity recognizer (BUFAR). Activity classification uses information of jaw movements (JM) within a segment. JM include: chew (c), bite (b), and chew-bite (cb).

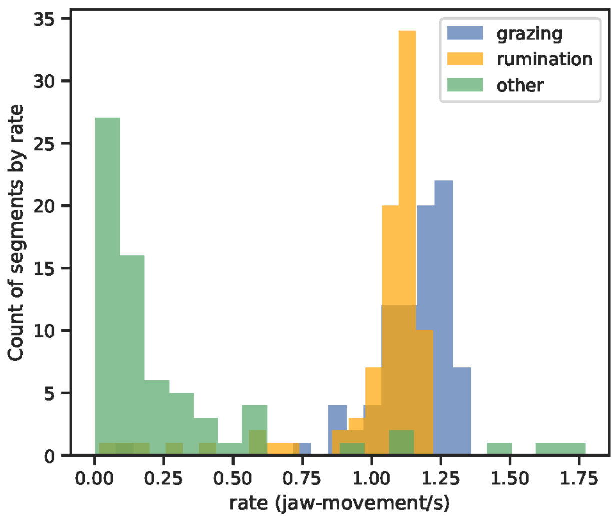

Figure 3: Distributions of jaw movements rate for grazing, rumination, and other activities on segments of the training set.

rate distributions of activities is part of the problem.

The triangle plot in Fig. 4 shows the proportions of the identified JM by CBIA for several segments of the training set. Proportions of a single segment %c, %b, and %cb always sum to 1.0. The top corner corresponds to 100% of chews, the bottom left corner corresponds to 100% of chew-bites, and the bottom right corner corresponds to 100% of bites. Points inside the triangle correspond to segments composed by more than one type of JM. For example, while rumination is mainly composed by chews (orange diamonds are on the top corner), grazing has a diversity of JM compositions (blue circles are dispersed in the triangle). During other activities, bites are the most assigned type of JM (green squares located on the bottom right corner). However, they are mostly false positives for class b.

Distributions of segment features show that the recognition of JM within

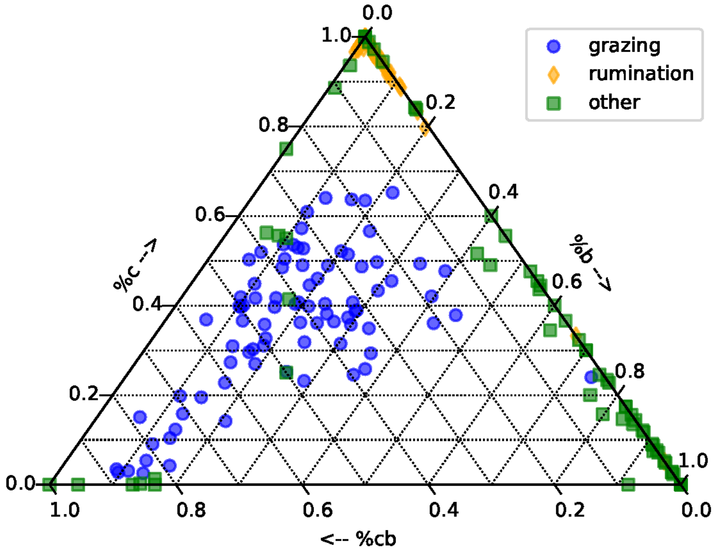

Figure 4: Proportions of bite (%b), chew (%c), and chew-bite (%cb) as labeled by Chew-Bite Intelligent Algorithm (CBIA) for grazing, rumination, and other activities. The top corner corresponds to 100% of chews, the bottom left corner corresponds to 100% of chew-bites, and the bottom right corner corresponds to 100% of bites.

grazing and rumination activities is not perfect. For example, CBIA detects a few bites during rumination, which is not actually true. Thus, the problem of distinguishing between activities requires a powerful method to handle these errors. In this study, the use of a simple method of machine learning is proposed. Activity classification is performed by a trainable model, such as a multilayer perceptron or a decision tree, which assigns an activity label to the segment. A multilayer perceptron (MLP) is a feed-forward artificial neural network that can deal with non-linearly separable data (Bishop, 2006). It consists of several layers of nodes (simple perceptrons) in a directed graph, with each layer fully connected to the next one, but without connections between nodes in the same layer. Decision Trees (DTs) have the ability of learning simple decision rules and systematizing them in order to arrive at complex decisions (Bishop, 2006). For numerical attributes, DTs divide the feature space into axis-parallel rectangular regions and label each region with the correspondent class. In addition, a DT provides solutions which are easy to implement and understand.

At the end of the processing stages, each segment of the input signal has a label that indicates if it corresponds to rumination, grazing, or other activity. Finally, a smoothing process is applied over the sequence of labeled segments in order to remove short gaps and thus reduce fragmentation of activity bouts. Thus, long recognized bouts are encouraged, which mimics the typical length of activity bouts.

2.2. Acoustic database

Acoustic signals were collected in August of 2014 at the dairy facility in the Kellogg Biological Station Robotic and Grazing Farm, operated by the

Michigan State University. As described in (Vanrell et al., 2018), the code for animal use by the Institutional Animal Care and Use Committee of the Michigan State University was reviewed, approved, and conducted according to protocols for animal handling and care. SONY ICDPX312 recorders were used to record the signals (Fig. 5a). A microphone was placed facing inwards on the forehead of cows (Fig. 5b) and was protected by a rubber foam (Milone et al., 2012). All recordings were saved in WAV file format, considering a 44.1 kHz sampling rate and 16-bit resolution.

Cows were rotationally grazed on a pasture-based robotic milking system with voluntary cow traffic as described previously in Watt et al. (2015). Briefly, the five multiparous experimental cows (parity 2.6 $\pm$ 0.5; days in milk 108 $\pm$ 34; body weight 654 $\pm$ 21 kg; milk yield 39 $\pm$ 4 kg; milkings/d 3 $\pm$ 1) were group housed and managed together as part of a larger robotic and grazing herd of 146 Holstein cows, allocated to two Lely A3-Robotic Milkers (Lely Industries N.V., Maassluis, the Netherlands). Cows were raised and grazed previously on same pasture so they were properly adapted to the farming system and diets before this study commenced. Milking was conducted according to milk table permissions set by a minimum expected milk yield/milking of 9.1 kg or 6 h of minimum interval. During milking cows were fed a grain based concentrate (GBC) at a rate of 1 kg per 6 kg of milk production (cap 12 kg/ cow d^{-1}). The average crude protein (CP), neutral detergent fiber (NDF), and net energy for lactation (NEL) of the GBC pellet offered (Cargill Inc, Big Lake, MN) was 193.0 g/kg DM, 99.4 g/kg DM, and 2.05 Mcal/kg DM, respectively. Cows had 24 h access to pasture dominated either by perennial ryegrass (Lolium perenne) and white clover

(Trifolium repens), or orchardgrass (Dactylis glomerata), tall fescue (Festuca arundinacea) and the same white clover. Cows were grazed at an average herbage allowance of 30 kg DM/cow d^{-1} split evenly into an AM and PM break of fresh pasture (15 kg DM/cow) freely accessible at opposite locations of the farm (north and south) from 10:00 h to 22:00 h and from 22:00 h to 10:00 h, respectively. Herbage allowance was adjusted according to changes in pasture growth rates and measurements of pregrazing herbage cover (Y; measured to ground level) by a plate meter ($Y = 125x$; $r^2 = 0.96$), using 30 readings of sward height (SH; x) taken alongside allocations. At the time of the study the average pregrazing and postgrazing herbage mass ($n = 16$ paddocks) was 2387 $\pm$ 302 kg DM/ha (19.2 $\pm$ 2.5 cm SH) and 1396 $\pm$ 281 kg DM/ha (11.2 $\pm$ 2.2 cm SH), respectively. The average CP (4010 CN combustion system, Costech Analytical Technologies Inc., Valencia, CA), NDF and acid detergent fiber (ADF) (200 Fiber Analyzer, Ankom Technology Corp., Fairport, NY), and acid detergent lignin (ADL) content and 48 h in vitro DM digestibility (Daisy II, Ankom Technology Corp.) of hand pluck pasture samples ($n = 16$) was 187 $\pm$ 25 g/kg DM, 493 $\pm$ 45 g/kg DM, 257 $\pm$ 20 g/kg DM, 33 $\pm$ 8 g/kg DM, and 78.1 $\pm$ 3.0%, respectively.

Expert labeling was used as a control reference for comparison and evaluation against algorithms results. Two experts with prior experience on animal behavior scouting, and digital analysis of acoustic signals, viewed the plot of the sound waveform and listened to the recordings to make a decision. Experts were able to identify, classify, and label the activity blocks, either as grazing, rumination, or neither of these activities. Experts agreed 100% on the labeling, but there were small differences in the limits of each label (start

Figure 5: Box located on the cow's neck: (a) microphone plug-in and (b) recording device. Internal view of (c) the headband and (d) the microphone protected by a rubber foam on the cow's forehead.

and end). These marks were carefully revised further until an agreement was made. This type of labeling was already used in previous works and validated with visual check, in-situ and using videos (Chelotti et al., 2018, 2016; Vanrell et al., 2018).

For comparison purposes on rumination time estimation, the animals were continuously monitored during the experiments with the Hi-Tag rumination monitor system. This system consists of rumination loggers on the collar of the animal, stationary or mobile readers, and software for processing electronic records (Schirmann et al., 2009). Animal behavior was monitored with this system and summarized as the total time spent ruminating during two-hour chunks.

2.3. Performance metrics

In continuous activity recognition, performance evaluation requires a comparison between a reference sequence and a recognized sequence. The activity blocks of the reference sequence and the recognized sequence may not be in a one-to-one correspondence. For example, a single block (an activity bout) of the reference sequence can be partially detected by three shorter blocks in the recognized sequence. A comprehensive set of performance metrics for continuous activity recognition has been proposed by Ward et al. (2011) and has been recently used in a related study (Vanrell et al., 2018). These metrics are based on two complementary short- and long-term timescales. They present a multidimensional and detailed description instead of a single performance number. In this way, the strengths and weaknesses of a recognizer can be assessed, avoiding ambiguity in the results. Short-term metrics are frame-based, which is a small fixed-length unit of time. Frame-based metrics facilitate a fine-grain analysis that resembles a continuous time analysis. By contrast, a block has no fixed-length and is defined as a continuous period of time of a sequence that has the same label. For example, a rumination block in the reference sequence is a rumination bout. Long-term metrics are block-based, which provide a different point of view, like a big picture of the recognition performance. This is particularly valuable to detect coarse-grain bias and to propose modifications in the recognizer.

The frame- and block-based error metrics were used to characterize each variant of the method. They are false negative rate (FNR_*), false discovery rate (FDR_*), recall (R_*), precision (P_*), fragmentation (F_*), merging (M_*), deletion (D_*), insertion (I_*), underfill (U_*), Overfill (O_*), and the standard

F1-score ($F1_*$). All metrics were computed for each recording analyzed and then averaged for results presentation. For details about the computation of these metrics see Appendix A.

2.4. Experimental Setup

In this study, the following setup was considered for the proposed method. Computer experiments were performed considering that at time t the algorithm can use data available at time t and $t - \Delta t$ but no using data at $t + \Delta t$. This consideration is equivalent to online processing within the device. The configuration of CBIA was the same used in Chelotti et al. (2018). For the signal pre-processing stage, a Least Mean Square filter was used (Widrow et al., 1975). This adaptive filter has proven to be useful for removing trends at low computational cost. For detection of JM, the steps proposed in Chelotti et al. (2018) were implemented. For classification of JM, it was selected a one-hidden-layer multilayer perceptron.

An exploratory analysis on a subset of the training set was conducted for the segment buffering stage. Segments of 1.0, 2.5, 5.0, and 10.0 min in length were considered. The shortest segment considered (1.0 min) can capture at least a typical period of rumination. In addition, this segment length generally includes a number of JM that allows a suitable analysis. Segments longer than 10.0 min would result in poor temporal resolution. For the activity classification stage, two models were considered: i) a multilayer perceptron (MLP) with one hidden layer and a logistic activation function, and ii) a binary decision tree (DT) based in the Gini impurity measure. An hyper-parameter optimization was performed for both activity classifiers considering: the number of neurons in the hidden layer and learning rate for the MLP, and the pruning factor for the DT. This optimization was made following a 5-fold scheme with signals on the other subset of the training set and maximizing the accuracy measure[1]. Finally, in the last stage, a smoothing process to avoid fragmentation in rumination and grazing bouts was applied: single segments were relabeled when they were surrounded by segments of the same activity.

For this study, 30 h of recordings containing rumination and grazing sessions were randomly selected to optimize the segment-length. Another set of 24 h of recordings were used to train an optimize parameters and hyper-parameters of the activity classifier and they were never used again. Classifiers were trained following a 5-fold scheme on the training set. Finally, the test results were obtained from a separate test set of 137 h of recordings, which were selected taking care that they correspond to a free-ranging environment. Those portions of the recordings captured inside the feeding barn were excluded from this study. The periods inside the feeding barn were identified acoustically by experts, guided by the environmental sound (machines, engines, and the reverberation inside the barn) and the distinctive sound of metal gates opening and closing, when the animals entered or left the barn. This selection has been guided by the labels (timestamps) provided by the experts and it is in agreement with the study that presents the RAFAR (Vanrell et al., 2018). The present work included a comparison with the RAFAR-MBBP variant.

A web demo of the method was developed with the tool (Stegmayer et al.,

[1]This stage was implemented in python using the scikit-learn package.

2016) and can be accessed at: https://sinc.unl.edu.ar/web-demo/bufar/.

3. Results

3.1. Segment-length effect

Table 1 shows the effect of segment length in activity recognition using an MLP as the activity classifier (BUFAR-MLP). Frame- and block-based F1-scores provide measures of the recognition in a short and long timescale, respectively. The shortest segment considered (1.0 min) achieved good frame-based metrics on grazing but very poor metrics on rumination. The longest segment considered (10.0 min) achieved good block-based metrics on grazing and poor metrics on rumination. A comparison of block-based metrics on grazing between 2.5-min and 5-min segments showed a notable improvement in favor of 5-min segments. Regarding rumination, a comparison between 2.5-min and 5-min segments showed remarkable improvements in frame- and block-based metrics for 5-min segments. Similar results were obtained using a DT as the activity classifier. In an overall assessment, 5-min segments achieved a strong performance for both frame- and block-based F1-score on the studied activities.

3.2. Activity classification

Two variants of BUFAR were evaluated: i) one using a decision tree as the activity classifier (BUFAR-DT) and ii) one using a multilayer perceptron as the activity classifier (BUFAR-MLP). In a previous study (Vanrell et al., 2018), RAFAR showed notable performance when the entire sound recording was available (offline analysis). It is the only known method that estimates

Table 1: F1-score metrics on activity classification for different segment lengths using Bottom-Up Foraging Activity Recognizer - Multilayer Perceptron (BUFAR-MLP).

	Grazing		Rumination	
Segment-length	Frame-based	Block-based	Frame-based	Block-based
1.0 min	0.849($\pm$0.161)	0.693($\pm$0.355)	0.516($\pm$0.340)	0.500($\pm$0.173)
2.5 min	**0.851**($\pm$0.165)	0.770($\pm$0.359)	0.631($\pm$0.311)	0.642($\pm$0.263)
5.0 min	0.812($\pm$0.181)	**0.822**($\pm$0.196)	**0.703**($\pm$0.274)	**0.743**($\pm$0.318)
10.0 min	0.764($\pm$0.314)	0.811($\pm$0.244)	0.611($\pm$0.336)	0.567($\pm$0.279)

both grazing and rumination bouts from acoustic signals. For comparison purposes, the RAFAR-MBBP variant was considered in this study (in the following referred as RAFAR). For a fair comparison between RAFAR and the proposed methods, the same limited data (5-min sound segments) was considered as the input.

A spider plot considering frame- and block-based metrics for grazing recognition is shown in Figure 6. A perfect recognizer would yield 0 for each error metric, which matches the boundary of the polygon. Frame-based metrics (gray side of the diagram) showed excellent FDR_f ($\sim$10%) and poor FNR_f (<40%) for both BUFAR variants. This means that most frames were correctly labeled as grazing, whereas some frames corresponding to grazing activity were not detected (false negatives). Deletions (D_f) and underfills (U_f) explain most of the undetected frames. The best FDR_f was achieved by BUFAR-MLP, while BUFAR-DT obtained a slightly lower FNR_f among variants. RAFAR presented the opposite situation, low FNR_f and high FDR_f. Regarding other metrics such as F_f, M_f, O_f, and I_f, the evalu-

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
J. O. Chelotti, S. R. Vanrell, L. Rau, J. Galli, A. M. Planisich, S.A Utsumi, D. H. Milone, L. Giovanini & H. L. Rufiner; "An online method for estimating grazing and rumination bouts using acoustic signals in grazing cattle"
Computers and Electronics in Agriculture, Vol. 173, pp. 105443, 2020.

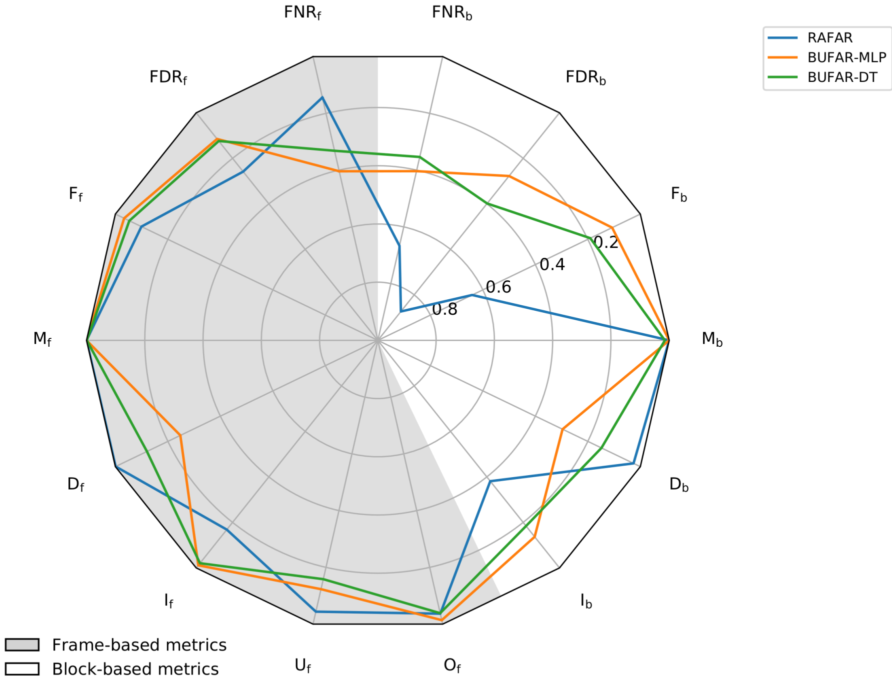

Figure 6: Spider plot of frame- and block-based metrics for grazing classification. Error metrics are: false negative rate (FNR), false discovery rate (FDR), fragmentation (F), merging (M), deletion (D), insertion (I), underfill (U) and overfill(O).The subscript indicates frame (f) or block-based (b) metrics.

ated variants achieved excellent results (<5%), which indicates that hardly any frame is associated with fragmentation, merging, overfill, or insertion of grazing.

Regarding the block-based analysis of grazing classification, BUFAR variants showed the lowest FDR_b and FNR_b and outperformed RAFAR on both metrics. BUFAR-MLP had slightly higher FNR_b but lower FDR_b than the BUFAR-DT. That is, BUFAR-MLP failed to detect some grazing block but added fewer extra grazing blocks (false positives) than BUFAR-DT. Both

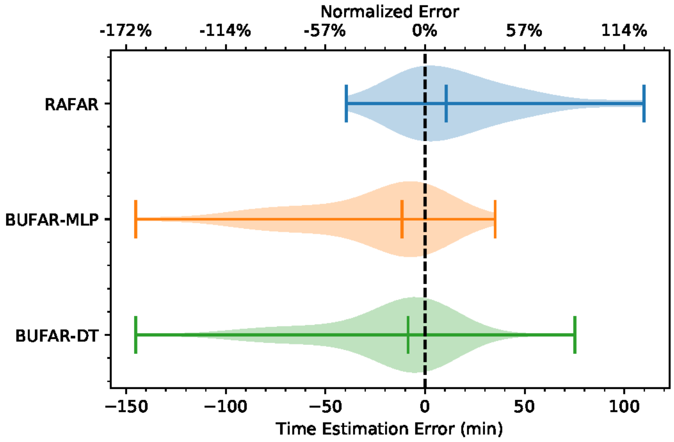

Figure 7: Violin plots of time estimation error for grazing classification. Normalized error considered the mean duration of grazing activity in the recordings (88.8 min).

variants achieved low fragmentation (F_b<20%) and insertion (I_b<20%), and very low merging (M_b<5%). These are great improvements over RAFAR performance on the same metrics.

The errors on duration estimation of grazing activity are shown in the violin plots of Figure 7. The distribution of the errors is shown for both variants and the RAFAR, across all recordings analyzed. The medians of the distributions show a tendency to underestimation for BUFAR-MLP and BUFAR-DT, while RAFAR overestimated grazing. Between BUFAR variants, BUFAR-MLP achieved a lower dispersion. Absolute median errors were below 12.0 min.

Recognition results of rumination activity are shown in Figures 8 and 9. In the spider plot of Fig. 8, proposed variants achieved better FDR_f and FNR_f metrics than RAFAR. BUFAR-MLP showed the lowest FNR_f, which makes it the variant that detected most of the actual rumination bouts.

Frames associated with fragmentation and merging of rumination bouts were very low and similar for both BUFAR variants. BUFAR-MLP achieved a notable lower underfill error compared with BUFAR-DT.

Regarding block-based results (white side of the diagram), rumination recognition showed similar FNR_b and FDR_b for BUFAR variants. Even though, there was a small difference in favor of BUFAR-MLP. These performance metrics were much better compared to the results obtained with the RAFAR. Results indicate that rumination blocks were rarely fragmented or deleted by the proposed method. In addition, hardly any rumination block was merged.

Finally, the time estimation error on rumination activity is shown in Figure 9. The lowest median was achieved by BUFAR-MLP (0.3 min). Also, BUFAR-MLP showed a lower dispersion than BUFAR-DT.

3.3. Overall performance

A summary of the evaluated methods is shown in Table 2. As a general performance indicator, the F1-score was computed for the RAFAR and both BUFAR variants. For this global measure, BUFAR variants clearly outperformed RAFAR for both grazing and rumination activities. This predominance is stronger on block-metrics, where 0.3 or higher improvements are seen. A comparison between the BUFAR variants showed similar results for grazing but a clear improvement for rumination in favor of BUFAR-MLP. Metrics differences between RAFAR and BUFAR variants has shown to be significant ($p<0.05$) using a Wilcoxon signed-rank test (Wilcoxon, 1945). Thus, BUFAR-MLP achieved the best and most consistent results in recognition among studied activities.

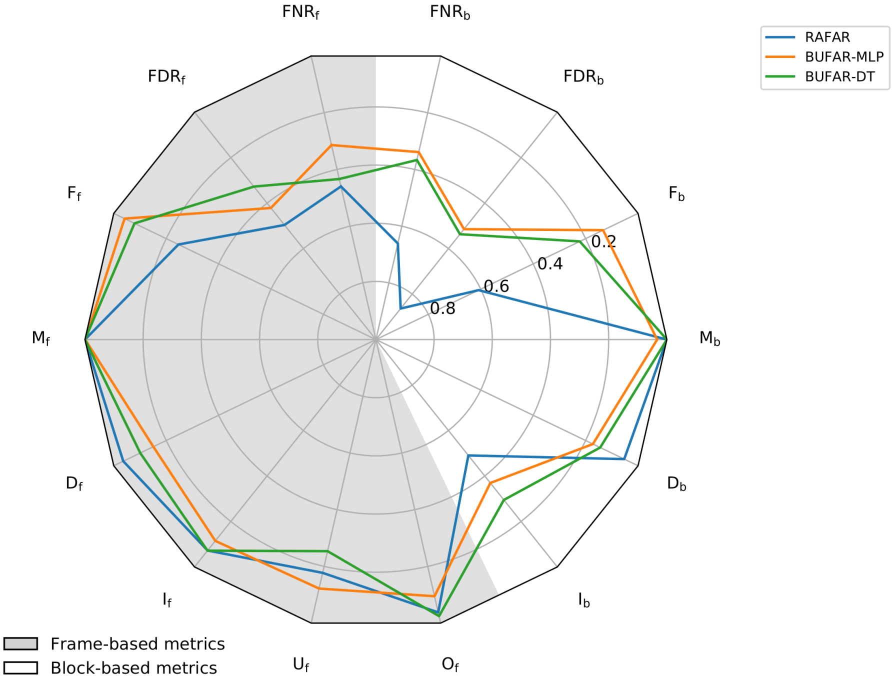

Figure 8: Spider plot of frame- and block-based metrics for rumination classification. Error metrics are: false negative rate (FNR), false discovery rate (FDR), fragmentation (F), merging (M), deletion (D), insertion (I), underfill (U) and overfill(O).The subscript indicates frame (f) or block-based (b) metrics.

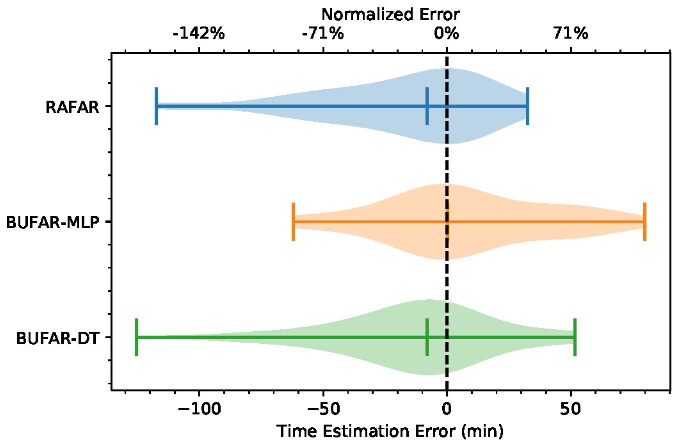

Figure 9: Time estimation error for rumination classification. Normalized error considered the mean duration of rumination activity in the recordings (69.9 min).

In order to evaluate the feasibility of online execution, the computational cost of the proposed method was computed (Table 2). Also, the computational cost of the former method RAFAR was calculated for the sake of comparison. Notice that the cost of the two BUFAR variants is the same because the impact of the classifiers is negligible compared to the other stages. The computational cost of these variants is about 50 times lower than the corresponding to RAFAR. A detailed description of computations for RAFAR and BUFAR are provided in Appendices B and C, respectively.

4. Discussion

The proposed online method BUFAR showed a good performance in the estimation of grazing and rumination bouts. Frame- and block-based measures provide different points of view of the recognition. While frame-based metrics provide a fine-grain analysis which approximates to continuous time,

Table 2: F1-score metrics on activity classification and computational cost (operations per second) of analyzed methods.

	Grazing		Rumination		Computational cost (ops/s)
	Frame-based	Block-based	Frame-based	Block-based	
RAFAR	0.783 (±0.180)	0.410 (±0.267)	0.633 (±0.240)	0.453 (±0.327)	1,892,354
BUFAR-MLP	**0.800** (±0.236)	**0.866** (±0.165)	**0.781** (±0.230)	**0.755** (±0.289)	**37,966**
BUFAR-DT	0.795 (±0.233)	0.819 (±0.229)	0.661 (±0.275)	0.734 (±0.303)	**37,966**

block-based metrics provide information about the recognition of activities as blocks providing a big picture view of the recognition. In particular, BUFAR-MLP achieved frame- and block-based F1-scores higher than 0.75 (Table 2) This consistency among metrics and activities made it the preferred variant of the proposed method.

4.1. Comparison with a former method

The block-based metrics achieved by BUFAR were much higher than the corresponding ones to RAFAR. That is, more actual activity bouts were correctly recognized as activity blocks. Regarding time estimation of activities, the absolute errors were low for BUFAR variants (medians below 12 min) compared to the mean duration of activities (Figures 7 and 9). No significant differences were observed on time errors between the RAFAR and proposed variants. The time estimation error is a practical but ambiguous performance metric. False negatives frames could be compensated by false positives frames. Thus, in this study the estimation error has been complemented with the frame- and block-based metrics. These considerations support that the performance achieved by BUFAR is meaningful and makes auspicious its implementation on a portable device.

Foraging activity recognition throughout online processing of the acoustic signal is a main goal in this study. That is, the proposed method must process data within the device. As a consequence, only monitoring results need to be stored in the device until they can be transferred to a central server in a farm. BUFAR follows this approach by analyzing the sound signal in real-time. JM are identified in the moment and an activity segment is defined every 5 min. On the contrary, a method as the RAFAR is meant to perform an offline processing, where an entire recording is required to obtain a proper result. The needs of massive volumes of data (several hours recordings) are not feasible for a limited device.

Another aspect to consider is the computational cost. Current microcontroller-based systems could operate at high frequency and perform heavy computations but at the expense of high power consumption. However, a method with low computational cost can be embedded in a microcontroller-based device working at low frequency and thus reducing the power consumption. This is essential for the development of a portable long-term monitoring device. The method proposed in this study requires 37,966 operations per second, which are much lesser than the 1,892,354 operations per second required by RAFAR. Thus, BUFAR is truly suitable to perform online processing.

The use of fixed-length segments minimizes computational cost. A segment is classified into an activity by computing only a few operations every few minutes (segment length), when the segment buffer has been filled with the detected JM. Thus, computational cost is not increased and the device requirements are not modified. The use of this kind of segments is a design choice. Actual duration of foraging-activity bouts is expected to be similar to a multiple of segment length but not exactly the same. Duration mismatches exist, which affect the performance of the system. An alternative to the use of fixed-length segments would be dynamic segmentation, i.e, the length of each segment would be determined adaptively according to the features of the sound signal. However, it is expected that a dynamic segmentation approach would significantly increase the computational cost, which goes against the goal of this study. An intermediate approach is to consider a Markov process, where each segment is independent when given the previous one (Milone et al., 2012). Both approaches could be explored in order to improve the recognition performance, considering its corresponding computational cost and online implementation.

4.2. Comparison with a commercial system

A comparison of the rumination time estimation obtained by the Hi-Tag system and the BUFAR-MLP was performed. The Hi-Tag system summarizes the total time the animal spent ruminating during two-hour chunks (Schirmann et al., 2009). Raw data and timestamps of rumination bouts within a two-hour chunk are not available (Goldhawk et al., 2013). Therefore, the estimations with the BUFAR-MLP were aligned, and total duration of rumination was summarized to match the same two-hour chunks of the Hi-Tag system. The comparison was made with a total of 53 two-hour chunks from all the recordings analyzed as it was done in (Vanrell et al., 2018). Due that the Hi-Tag is a commercial system, its computational cost was not

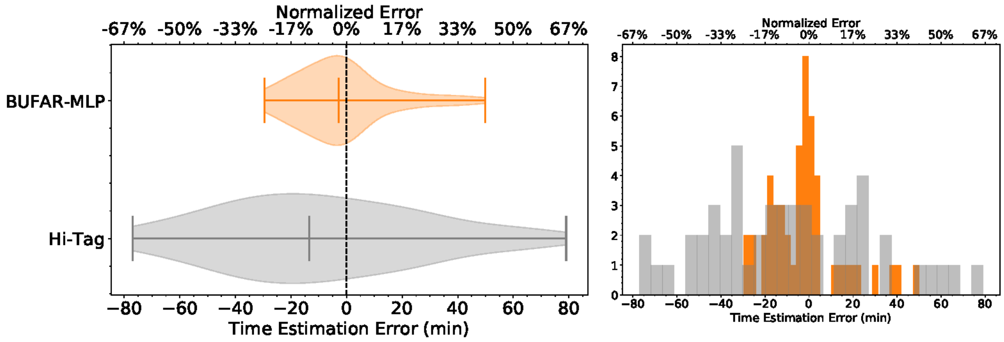

Figure 10: Time estimation error of rumination for BUFAR-MLP (orange) and Hi-Tag (gray). Top axis is normalized with the length of segments analyzed (2 hours).

available to be considered in the analysis.

The results of time estimation error for rumination are shown in Fig.10. The medians of the distributions are -2.91 min and -13.55 min for BUFAR-MLP and the Hi-Tag system, respectively. Negative medians imply that both systems tend to underestimate the rumination time. The underestimation shown by the Hi-Tag agree with previous studies that evaluate its performance (Burfeind et al., 2011; Goldhawk et al., 2013). BUFAR-MLP was more accurate and resulted in a narrower error distribution. While the error dispersion for BUFAR-MLP is in the range (-30, +50) min, the distribution corresponding to the Hi-Tag is wider and it is in the range (-80, +80) min. In practical terms, these errors are very high since they are in the same order of magnitude of the two-hour chunks analyzed. The wide dispersion shown by the Hi-Tag has been seen in previous studies (Burfeind et al., 2011; Goldhawk et al., 2013).

5. Conclusions

In this study, an online method for recognition and estimation of foraging activity bouts from acoustic signals has been presented. The proposed method BUFAR follows a bottom-up approach, which goes from jaw movement recognition to foraging activity recognition. Sound signals are processed and downsampled to operate at a lower frequency, aiming at the implementation of the method in a microcontroller-based system with limited resources. The recognition of grazing and rumination bouts was evaluated with specific metrics for activity recognition. Analyzing the results, the preferred variant of the proposed method is the BUFAR-MLP and medium-length segments. In addition, the BUFAR-MLP was superior in comparison with the former method RAFAR. Another important advantage is that the proposed method performs very few operations to recognize activity bouts. This ease the possibility of an online implementation for its execution on a low-cost embedded system. An additional comparison showed that the proposed method outperformed the Hi-Tag commercial system on rumination time estimation. Thus, the BUFAR good performance and simplicity achieved the stated goals. Future works could be focused on improving the recognition performance by including more complex features or processing techniques at the expense of an increased computational cost.

Acknowledgments

This study 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 study was funded under the USDA-NIFA MICL0222 and MICL0406 projects, and direct support from AgBioResearch-MSU. Authors wish to thank the insightful help and dedication by the KBS Robotic Dairy Farm staff for their assistance and support during the completion of this study.

Appendix A. Definitions of frame- and block-based error metrics

The frame- and block-based error metrics are defined in Table A.1. Frame-based metrics are defined by considering the counts of true positives TP, false positives FP, false negatives FN, fragmented F, merged M, deleted D, and underfill U frames in the reference sequence, and by the count of inserted I, and overfill O frames in the recognized sequence, respectively. Frames of 1 s were considered as the smallest time unit for analysis. Block-based metrics are defined by considering the counts of total (B^{ref}), correctly detected (C), fragmented (F), merged (M), and deleted (D) blocks in the reference sequence, and by the counts of total (B^{rec}) and inserted (I) blocks in the recognized sequence, respectively. In addition, the standard F1-score was computed for frames $F1_f = \frac{2R_fP_f}{R_f+P_f}$ and blocks $F1_b = \frac{2R_bP_b}{R_b+P_b}$ based on the corresponding precision and recall defined in Table A.1.

Appendix B. Computational cost of RAFAR

The computational cost of RAFAR-MBBP (Vanrell et al., 2018) depends on the sampling frequency (S_f) and duration (T) of the input signal. In

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

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

order to get a straightforward comparison with other algorithms, a sampling frequency of $S_f = 2$ kHz and a duration of $T = 300$ s were selected to compute the computational cost. 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 RAFAR-MBBP was:

1. **Segmentation by regularity**
   (a) *Envelope computation*: This task comprise signal rectification, signal filtering, and signal subsampling. First, signal rectification requires a comparison and a multiplication per sample. Second, a 3rd-order IIR low-pass filter is applied, which involves 7 multiplications and 6 additions per sample. Third, the envelope is sub-sampled at 1 kHz, which requires 1,000 comparisons/s.
   (b) *Regularity analysis*: The envelope is analyzed by frames of 30 s. The computation of autocorrelation requires $29.225 \cdot (S_f/2) \cdot 951$

multiplications and $[29.225 \cdot (S_f/2) - 1] \cdot 951$ additions for each 30 s frame. Then, a peak is searched, which requires 12,264 comparisons for each 30 s frame. Once a peak is found, the regularity rule is evaluated with two comparisons and the frame is labeled in one assignment.

   (c) *Smoothing filter*: A 5th-order median filter is implemented, which involves 10 comparisons for each 30 s frame.

   The computational cost of the segmentation stage is 565,272,760 operations.

2. **Classification of activity blocks**

   (a) *Energy computation*: This task is performed using 1 s frames and requires $2 \cdot S_f + 4$ multiplications, $2 \cdot S_f + 2$ additions, and 3 assignments per frame.

   (b) *Sudden-drop detection*: Worst case scenario considers an 80 s sliding window with a 5 s step. The median of the energy is computed with 507 comparisons per window. A threshold is generated and compared requiring 1 multiplication and 1 comparison per window.

   (c) *Rules classification*: This task required 4 comparisons for each activity block.

   The computational cost of this stage is 1,233,244 operations.

3. **Block partition**: Worst-case scenario for this stage is to consider the input signal as a single block. Computation of block duration requires 1 subtraction. A block is analyzed if the duration is greater than 10 min, which requires 1 comparison. A block is analyzed with 60 s frames.

Energy is computed requiring $2 \cdot 60 \cdot S_f + 4$ multiplications, $2 \cdot 60 \cdot S_f + 2$ additions, and 3 assignments per frame. The detection of changes in the computed energy requires 1 multiplication and 1 comparison with a threshold, for each 60 s frames. If a block should be partitioned, 2 extra assignments are required. Therefore, the computational cost of this stage is 1,200,059 operations.

4. **Merging gaps**: The worst-case scenario for this stage is to consider that the entire input signal has the shortest activity blocks and the shortest inactivity gaps. A subtraction is required to compute the duration of the gap and it is compared with a threshold. If a gap should be merged, 3 extra assignments are required. Therefore, the computational cost of this stage is 9 operations.

The overall computational cost for the RAFAR-MBBP is: $565{,}272{,}760 + 1{,}233{,}244 + 1{,}200{,}059 + 9 = 567{,}706{,}072$ operations. The most computational-expensive stage is the segmentation, which requires 99.57% of the total operations. Specifically, the autocorrelation computation requires 97.91% of the total operations.

To compare the RAFAR-MBBP with an online method and considering the duration of the input signal (300 s), the computational cost can be estimated as 1,892,354 operations/s.

Appendix C. Computational cost of BUFAR

The computational cost of BUFAR depends on the sampling frequency (fixed in 2 kHz in this analysis) and the duration (fixed in $T = 300$ s in this analysis) of the input signal. A 5 min segment length and 2 jaw movements per second were selected in order to consider the worst-case scenario in the sense of computational cost. The required number of operations per second for the computation stages of BUFAR was:

1. **Signal pre-processing**: A least mean square filter (LMS) requires 5 operations per signal sample. Then, 10,000 operations/s are required.
2. **Jaw-movement detection**: 27,800 operations/s are required to detect jaw movements and to extract their features.
3. **Jaw-movement classification**: MLP requires 80 operations per jaw movement, thus, 160 operations/s are required.
4. **Segment buffering**: this stage requires 6 operations/s and 6 operations per segment to save the timestamp and to compute the segment features.
5. **Activity classification**: this stage was evaluated for MLP and DT. MLP requires 170 operations per segment. DT requires 6 operations per segment.
6. **Smoothing process**: to avoid fragmentation in rumination and grazing bouts, 2 comparisons per segment are required.

Hence, the overall computational cost is 37,966 operations/s + 178 operations/segment for BUFAR-MLP, and 37,966 operations/s + 14 operations/segment for BUFAR-DT. The costs of activity classification and smoothing process are negligible because the operations are performed just a few times in a long period of time (segment length).

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. Biotechnology, Agronomy, Society and Environment 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.

Arcidiacono, C., Porto, S.M.C., Mancino, M., Cascone, G., 2017. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data. Comput. Electron. Agric. 134, 124–134.

Beauchemin, K.A., 1991. Ingestion and mastication of feed by dairy cattle. Vet. Clin. North Am. Food Anim. Pract. 7, 439–463.

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

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

Bristow, D.J., Holmes, D.S., 2007. Cortisol levels and anxiety-related behaviors in cattle. Physiol. Behav. 90, 626–628.

Burfeind, O., Schirmann, K., von Keyserlingk, M., Veira, D., Weary, D., Heuwieser, W., 2011. Evaluation of a system for monitoring rumination in heifers and calves. Journal of dairy science 94, 426–430.

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

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

Clapham, W.M., Fedders, J.M., Beeman, K., Neel, J.P.S., 2011. Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle. Comput. Electron. Agric. 76, 96–104.

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.

Giovanetti, V., Decandia, M., Molle, G., Acciaro, M., Mameli, M., Cabiddu, A., Cossu, R., Serra, M.G., Manca, C., Rassu, S.P.G., Dimauro, C., 2017. Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer. Livest. Sci. 196, 42–48.

Goldhawk, C., Schwartzkopf-Genswein, K., Beauchemin, K.A., 2013. Validation of rumination collars for beef cattle. J. Anim. Sci. 91, 2858–2862.

González, L.A., Bishop-Hurley, G.J., Handcock, R.N., Crossman, C., 2015. Behavioral classification of data from collars containing motion sensors in grazing cattle. Comput. Electron. Agric. 110, 91–102.

Greenwood, P.L., Paull, D.R., McNally, J., Kalinowski, T., Ebert, D., Little, B., Smith, D.V., Rahman, A., Valencia, P., Ingham, A.B., Bishop-Hurley, G.J., 2017. Use of sensor-determined behaviours to develop algorithms for pasture intake by individual grazing cattle. Crop Pasture Sci. .

Hansen, S.S., Nørgaard, P., Pedersen, C., Jørgensen, R.J., Mellau, L.S.B., Enemark, J.D., 2003. The effect of subclinical hypocalcaemia induced by Na2EDTA on the feed intake and chewing activity of dairy cows. Vet. Res. Commun. 27, 193–205.

Herskin, M.S., Munksgaard, L., Ladewig, J., 2004. Effects of acute stressors on nociception, adrenocortical responses and behavior of dairy cows. Physiol. Behav. 83, 411–420.

Hodgson, J., Illius, A.W., 1998. The ecology and management of grazing systems. Wallingford (United Kingdom) CAB International.

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

Kilgour, R.J., 2012. In pursuit of "normal": A review of the behaviour of cattle at pasture. Appl. Anim. Behav. Sci. 138, 1–11.

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, 81–90.

Martiskainen, P., Järvinen, M., Skön, J.P., Tiirikainen, J., Kolehmainen, M., Mononen, J., 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Appl. Anim. Behav. Sci. 119, 32–38.

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.

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, 474–483.

Nydegger, F., Gyga, L., Egli, W., 2010. Automatic measurement of rumination and feeding activity using a pressure sensor, in: International Conference on Agricultural Engineering - AgEng 2010.

Paudyal, S., Maunsell, F.P., Richeson, J.T., Risco, C.A., Donovan, D.A., Pinedo, P.J., 2018. Rumination time and monitoring of health disorders during early lactation. Animal 12, 1484–1492.

Phillips, C.J.C., 1993. Cattle Behaviour. Farming Press.

Rutter, S.M., 2000. Graze: a program to analyze recordings of the jaw movements of ruminants. Behav. Res. Methods Instrum. Comput. 32, 86–92.

Rutter, S.M., Champion, R.A., Penning, P.D., 1997. An automatic system to record foraging behaviour in free-ranging ruminants. Appl. Anim. Behav. Sci. 54, 185–195.

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.

Stegmayer, G., Pividori, M., Milone, D.H., 2016. A very simple and fast way to access and validate algorithms in reproducible research. Brief. Bioinform. 17, 180–183.

Ungar, E.D., Rutter, S.M., 2006. Classifying cattle jaw movements: Comparing IGER behaviour recorder and acoustic techniques. Appl. Anim. Behav. Sci. 98, 11–27.

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. Comput. Electron. Agric. 151, 392–402.

Ward, J.A., Lukowicz, P., Gellersen, H.W., 2011. Performance metrics for activity recognition. ACM Trans. Intell. Syst. Technol. 2, 6:1–6:23.

Watt, L., Clark, C., Krebs, G., Petzel, C., Nielsen, S., Utsumi, S., 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.

Welch, J.G., 1982. Rumination, particle size and passage from the rumen. J. Anim. Sci. 54, 885.

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. J. Neurosci. Methods 300, 138–146.

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, 1692–1716.

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

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. Comput. Electron. Agric. 136, 31–41.