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

Sebastián R. Vanrell[a], José O. Chelotti[a], Julio R. Galli[b], Santiago A. Utsumi[c], Leonardo L. Giovanini[a], H. Leonardo Rufiner[a,d], Diego H. Milone[a]

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

Abstract

Continuous monitoring of cattle foraging behavior is a major requirement for precision livestock farming applications. Several strategies have been proposed for this task but monitoring of free-ranging cattle for a long period of time has not been fully achieved yet. In this study, an algorithm is proposed for long-term analysis of foraging behavior that uses the regularity of this behavior to recognize grazing and rumination bouts. Acoustic signals are analyzed offline in two main stages: segmentation and classification. In segmentation, a complete recording is analyzed to detect regular masticatory events and to define the time boundaries of foraging activity blocks. This stage also defines blocks that correspond to no foraging activity (resting bouts). The detection of event regularity is based on the autocorrelation of the sound envelope. For classification, the energy of sound signals within a block is analyzed to detect pauses and to characterize their regularity. Rumination blocks present regular pauses, whereas grazing blocks do not. The evaluation of the proposed algorithm showed very good results for the segmentation task and activity classification. Both tasks were extensively analyzed with a new set of multidimensional metrics. Frame-based F1-score was up to 0.962, 0.891 and 0.935 for segmentation, rumination classification,

and grazing classification, respectively. The average time estimation error was below 0.5 min for classification of rumination and grazing on recordings of several hours in length. In addition, a comparison for rumination time estimation was done between the proposed system and a commercial one (Hi-Tag; SCR Engineers Ltd., Netanya, Israel). The proposed algorithm showed a narrower error distribution, with a median of -2.56 min compared to -13.55 min in the commercial system. These results suggest that the proposed system can be used in practical applications.
Web demo available at: http://sinc.unl.edu.ar/web-demo/rafar/

Keywords: acoustic monitoring, activity recognition, grazing cattle behavior, precision livestock farming, signal processing

1. Introduction

In recent years, much effort has been put into the development of animal monitoring applications for precision livestock farming. Monitoring of foraging behavior is key to ensure the fulfillment of the basic health and welfare requirements of grazing cattle and to improve the efficiency of pasture-based production systems (Hodgson and Illius, 1998). Foraging activities, particularly grazing and rumination, occupy most of the animal's day. Thus, the continuous monitoring of such behavior can help retrieve individual status information for each animal, build a log, detect emerging diseases or the onset of estrus, and optimize pasture and animal management. For example, decreased rumination is interpreted as an indicator of stress (Herskin et al., 2004), anxiety (Bristow and Holmes, 2007), or disease (Welch, 1982). Conversely, an increase in rumination time is associated with more saliva production and improved rumen health (Beauchemin, 1991).

Cattle foraging behavior is mainly composed of grazing and rumination times. Grazing can cover from 25% to 50% of the day and rumination, from 15% to 40% (Kilgour, 2012). The grazing process involves searching, apprehending, chewing, and swallowing herbage. Rumination includes bolus regurgitation, chewing, and deglutition. While grazing, the animal moves its jaw continuously with no predefined interruptions or sequence of events. By contrast, a typical rumination phase involves chewing for 40-60 seconds and a

Email address: `srvanrell@sinc.unl.edu.ar` (Sebastián R. Vanrell)

3-to-5-second interruption during bolus deglutition and regurgitation (Hodgson and Illius, 1998; Trindade et al., 2011; Benvenutti et al., 2016). During both activities, jaw movements (or masticatory events) are performed rhythmically with a frequency that ranges from 0.75 to 1.20 events per second (Andriamandroso et al., 2016). The masticatory events are biting, when herbage is apprehended and severed; chewing, when herbage is comminuted; and a compound movement called chew-bite, when herbage is severed and comminuted in the same jaw movement (Laca et al., 1992; Ungar and Rutter, 2006; Galli et al., 2017). Events have a length close to 1 s, whereas activity bouts can last from minutes to hours. Thus, foraging behavior is characterized by events (short timescale) and activities (longer timescale).

Many strategies have been proposed for monitoring foraging behavior, but they are limited by several factors (Andriamandroso et al., 2016; Delagarde et al., 1999; Hodgson and Illius, 1998). For instance, foraging behavior could be measured by direct observation or by watching video recordings. However, these methodologies are extremely time-consuming and unfeasible for large herds; besides, it is very difficult to collect data in pasture-based systems over long periods of time. To be of practical use, monitoring should be performed in a fully automatic and noninvasive manner so as not to disturb the normal behavior of the animal. In addition, the system should be capable of working continuously and keep accurate measurements from days to weeks.

Automatic monitoring systems have been developed based on different sensing technologies: motion sensors, noseband pressure sensors, and microphones. The most commonly used motion sensors are accelerometers (González et al., 2015; Arcidiacono et al., 2017; Giovanetti et al., 2017; Martiskainen et al., 2009) and inertial measurement units (Andriamandroso et al., 2017; Smith et al., 2016; Greenwood et al., 2018). These systems typically seek to recognize a broader set of activities, such as rumination, grazing, resting, drinking, and walking. An activity is determined by postural analysis of the animal, where the sensors are used to estimate the relative position and motion of its head and body. However, this strategy can confuse activities that share the same posture. For example, resting can be easily confused with rumination, which can be performed while the cow is standing or lying on the ground. A better strategy for recognizing ruminating, eating, and drinking activities is the use of noseband pressure sensors (Rutter et al., 1997; Rutter, 2000; Nydegger et al., 2010; Zehner et al., 2017; Werner et al., 2018). The IGER Behavior Recorder was a pioneer development using these sensors. Recently, the RumiWatch system was used to analyze housed and free-ranging cows during one- and two-hour sessions. This yielded very good results, but further studies are required on continuous long-term monitoring. By contrast, acoustic monitoring has proven to be reliable for recognizing short-term ingestive events in free-ranging cows (Laca et al., 1992; Galli et al., 2011; Clapham et al., 2011; Navon et al., 2013; Milone et al., 2012; Galli et al., 2017; Chelotti et al., 2016, 2018). A popular monitoring system that includes a logger with a built-in microphone is the Hi-Tag system (SCR Engineers Ltd., Netanya, Israel). However, the sound signal processing is exclusively focused on monitoring rumination in housed cows (Schirmann et al., 2009; Goldhawk et al., 2013). No long-term acoustic monitoring of foraging activities has yet been studied for free-ranging cows.

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

The identification of rumination and grazing bouts can be seen as a particular case of continuous activity recognition problem. In this context, recognition systems are typically assessed with standard performance metrics, such as sensitivity, specificity, precision, or correlation coefficient (concordance, Pearson, or Spearman) (Sokolova and Lapalme, 2009; Werner et al., 2018; Zehner et al., 2017). However, to use these metrics the problem of continuous activity recognition must be reformulated as a classic classification problem, where input data is mapped to a single category. Unfortunately, restating the problem to conform to standard metrics can be misleading and can produce confusing results (Ward et al., 2011). In this study, we propose the use of a new set of multidimensional performance metrics, which provides a detailed description of the recognition process at multiple timescales. This allows for a more accurate assessment of the strengths and weaknesses of the proposed recognizers.

2. Materials and methods

Grazing and rumination are activities with quasiperiodic characteristics. The proposed *regularity-based algorithm* aims to use this discriminative information to provide grazing and rumination bouts. Two main stages are involved in the offline recognition process: activity segmentation and activity classification (Figure 1). The complete recording is first analyzed to delimit the blocks of the signal that show regular events (jaw movements). A short sliding window on the envelope of the sound signal is used to analyze this regularity. Demarcation of the activity blocks also defines blocks of no activity (resting bouts), which correspond to silence or noisy intervals. Autocorrelation is a well-known technique that has been useful to detect periodicity in noisy signals (Oppenheim and Schafer, 2011) and it will be used in this stage. During classification, activity blocks are further analyzed to detect interruptions and to characterize their regularity. The energy of the sound signal within a block is analyzed to detect sudden drops, which are related to the interruptions. Regular interruptions are related to bolus deglutition and regurgitation during rumination. Grazing does not show interruptions corresponding to this particular regularity, although it may present irregular interruptions by searching a new plant or patch.

2.1. Segmentation by regularity

Segmentation is based on regularity of masticatory events during grazing and rumination. The analysis of the envelope of the sound signal can reveal these events and their periodicity. Envelope computation is the first task of this stage (Figure 1.a). It allows one to operate with low-frequency signals and to discard high-frequency details, unrelated to event regularity (Chelotti et al., 2016). Envelope computation requires three steps: (i) signal rectification, (ii) signal filtering, and (iii) signal subsampling. In the first step the absolute value of signal samples is computed. In the second step the signal is filtered using a low-pass filter, thereby producing the sound envelope. In the third step, a subsample of the original sound envelope is conducted. The main objective of this step is to reduce the computational requirements in the subsequent tasks, since this process significantly reduces the amount of information to be processed without compromising the performance of the algorithm.

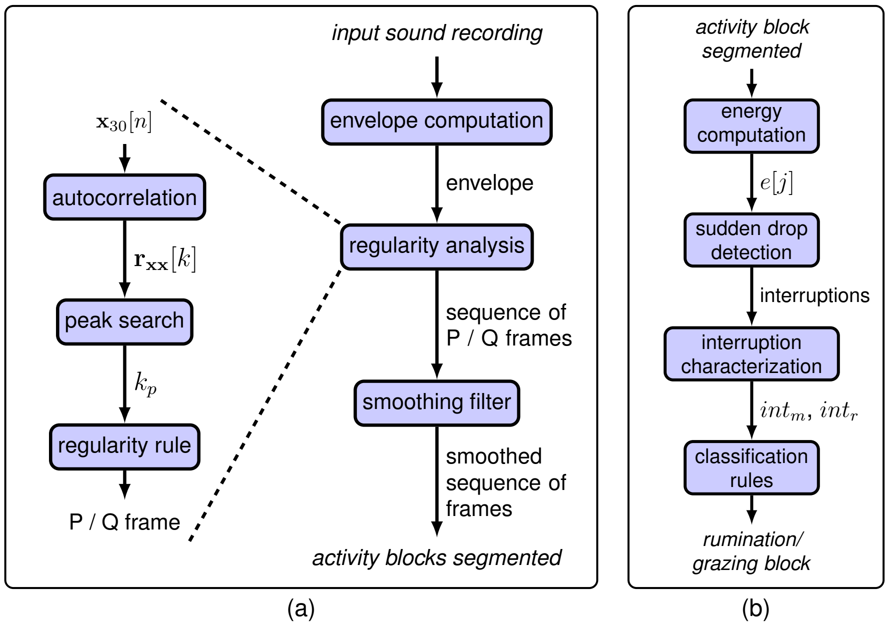


Figure 1: Tasks of a) segmentation and b) classification stages for the proposed algorithm. Steps of regularity analysis during segmentation are also detailed.

Sound envelope is analyzed by frames of 30 s without overlapping (Figure 1.a). The autocorrelation of each frame $x_{30}[n]$ is performed,

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

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

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

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

A sequence of labeled frames (as P or Q) is obtained after the whole recording has been analyzed (Figure 1.a). A few intermediate Q frames are

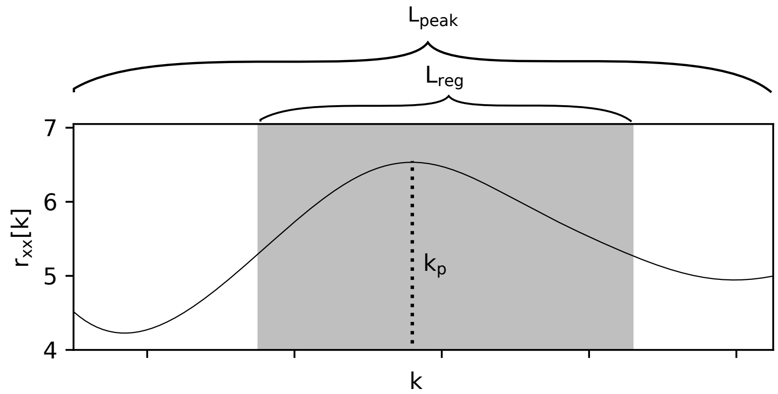


Figure 2: Autocorrelation $r_{xx}[k]$ of a frame $x_{30}[n]$ identified as regular. Maximum search in L_{peak} yielded k_p in L_{reg}.

undesired breaks from the activity point of view because they lead to fragmentation and unrealistic short activity blocks in the recognized sequence. Since activity segmentation is focused on long-term behavior, these breaks should be reduced. A smoothing filter is applied to labeled sequences to avoid the presence of one or two successive Q frames (<1 min) surrounded by P frames. Activity classification is more reliable when long blocks (several minutes) are analyzed.

2.2. Classification of activity blocks

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

The sound in an activity block is analyzed by 1 s frames $x_1[n]$ to detect interruptions related to bolus regurgitation (pointed by arrows in Figure 3.b). These brief interruptions (3-to-5 s) might be undetected using longer frames, thus 1 s frames were chosen for the analysis. Interruptions could be inferred

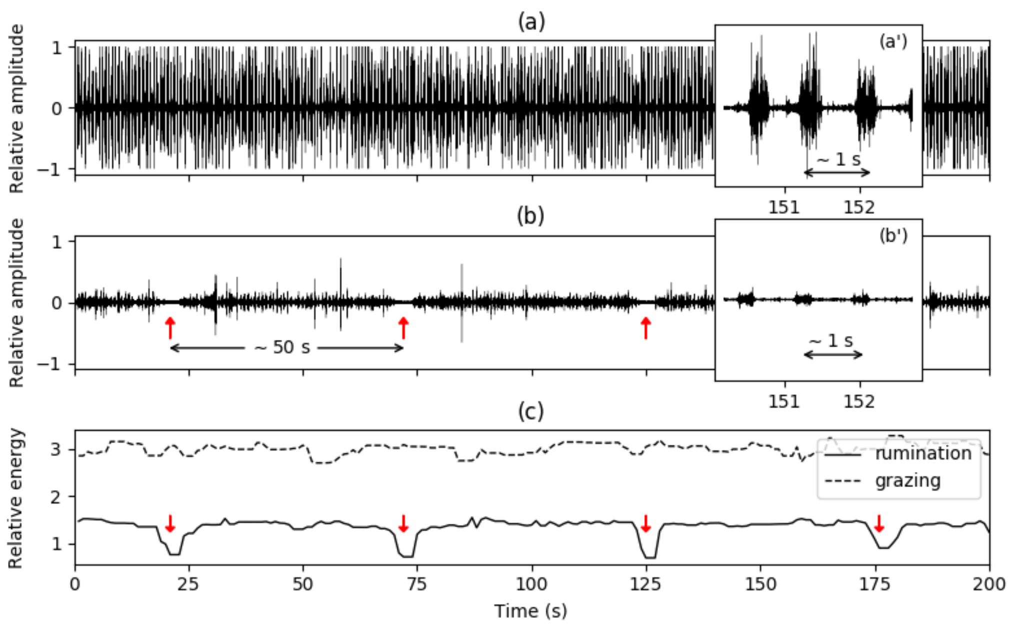


Figure 3: Sound recorded during a) grazing and b) rumination in a free-ranging environment. c) Energy of sound wave during grazing (dashed line), and rumination (solid line). Interruptions related to bolus regurgitation are indicated by arrows. Regularity of events is depicted in a') and b').

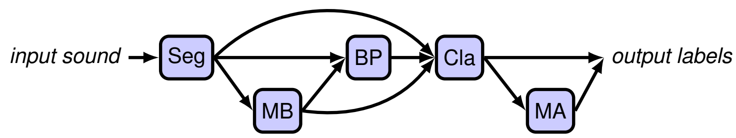
from changes in the amplitude of the sound signal. However, variations in the amplitude across sessions, microphones, recording devices and cows do not allow a reliable detection. Preliminary experiments showed that an energy-related measure performs better at detecting interruptions. Hence, in this study a frame is characterized by computing the $e[j] = \log(\|\mathbf{x}_1\|_2)$, which is proportional to the energy (Figure 1.b). The center of a frame j indicates its position in that block. For example, the solid line in Figure 3.c shows the corresponding $e[j]$ for the sound recorded during rumination (Figure 3.b). The algorithm aims to detect sudden drops in $e[j]$ so as to characterize interruptions.

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

2.3. Processing blocks and gaps

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

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



RAFAR:	*input sound* → Seg → Cla → *output labels*
RAFAR-MB:	*input sound* → Seg → MB → Cla → *output labels*
RAFAR-MA:	*input sound* → Seg → Cla → MA → *output labels*
RAFAR-BP:	*input sound* → Seg → BP → Cla → *output labels*
RAFAR-BPMA:	*input sound* → Seg → BP → Cla → MA → *output labels*
RAFAR-MBBP:	*input sound* → Seg → MB → BP → Cla → *output labels*

classification, gap merging is performed only if previous and posterior blocks correspond to the same activity. In this case, short gaps are combined with contiguous blocks to form a long block that is equally labeled. For instance, a short gap between two rumination blocks will be reassigned and a single longer rumination block will be created.

In this study, several variants are evaluated in the combination of these stages of the algorithm. They differ in the order of stages in the process flow. Their names and stages are detailed in Table 1. Available stages are segmentation by regularity (Seg), gap merging before classification (MB), partition of long blocks (BP), classification of activity blocks (Cla), and gap merging after classification (MA). The baseline variant is called regularity-based acoustic foraging activity recognizer (RAFAR). Merging stages are mutually exclusive, thus each variant will include MB or MA. The reason is that a gap that would be merged by MA will have already been merged by MB. A detailed description of the parameters of these stages is given in Section 2.5.

2.4. Acoustic signal database

The acoustic signals were obtained from an experiment performed at the dairy facility in the Kellogg Biological Station (Michigan State University), in August 2014. Protocols for animal handling and care were reviewed, approved, and conducted according to the Institutional Animal Care and Use Committee of Michigan State University. In this experiment, the for-

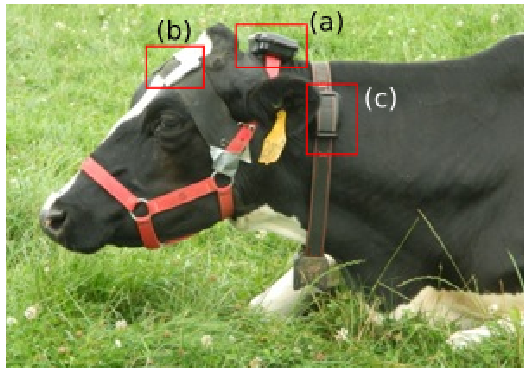


Figure 4: (a) Recording device, (b) microphone and (c) Hi-Tag logger location on the cow's head and neck.

aging behavior of five multiparous Holstein lactating cows grazing perennial ryegrass/white clover and orchardgrass/white clover pastures was continuously monitored during six days. These signals were recorded using SONY ICDPX312 recorders (Figure 4.a). A microphone was placed facing inwards on the forehead of cows (Figure 4.b) and was protected with rubber foam (Milone et al., 2012). All recordings were made at 44.1 kHz sampling rate and 16-bit resolution, providing a nominal 22 kHz recording bandwidth and 96 dB dynamic range, and they were saved in WAV (waveform audio) file format. For this study, 24 hours of recordings containing rumination and grazing sessions were selected to tune the parameters of the algorithm and they were never used again. The results were obtained from 137 hours of recordings, which were selected taking care that they corresponded to a free-ranging environment. Those portions of the recordings that were captured inside the feeding barn were excluded from this study. This selection has been guided by the labels (time-stamps) provided by the experts.

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

For comparison purposes on rumination time estimation, the Hi-Tag rumination monitor system was used to continuously monitor the animals during the experiments. The Hi-Tag system, consists of rumination loggers, stationary or mobile readers, and software for processing electronic records (Schirmann et al., 2009). Rumination was recorded with this system using a built-in microphone on the collar of the animal (Figure 4.c) and it was summarized as the total time spent ruminating during two-hour chunks.

2.5. Experimental setup

The experiments conducted considered the following implementation of the regularity-based algorithm. During segmentation, in the second step of the envelope computation, the signal was filtered using a third-order low-pass Butterworth filter with a cutoff frequency of 2 Hz. In the third step, a subsample of the original sound envelope to 1 kHz was conducted. Sound envelope was analyzed by frames of 30 s and the autocorrelation of each frame $x_{30}[n]$ was computed. This length provided enough time resolution to catch the periodicity peak in the autocorrelation and the desired time resolution for segmentation. Regular activities were expected to have a peak at around 0.8 s. This is related to the typical frequency of masticatory events, which is slightly higher than 1 Hz (Andriamandroso et al., 2016). The local maximum was searched in a surrounding interval L_{peak}, which corresponded to (0.3 s, 1.25 s). The frame was considered positive if the maximum was in $L_{\text{reg}} \subset L_{\text{peak}}$. This interval corresponded to the interval (0.55 s, 1.06 s) that covered the typical period of masticatory events. The smoothing filter applied to the labeled sequences was a fifth-order median filter. The intervals were defined from experiments with the 24 hour of recordings detailed in Section 2.4. The remaining parameters was defined from preliminary experiments with signals similar to those used in this study.

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

In the block partition stage, larger blocks (10 min or longer) were analyzed for possible partition. Frames of 60 s were considered to compute the energy and the partition was performed if the relative energy change was greater than a threshold of 0.4. A partition was not performed when the resulting blocks were shorter than 200 s. In the merging stages, a 5 min or shorter gap was considered a short gap that should be merged. These parameters were fixed from experiments with the 24 hour of recordings detailed in Section 2.4. A web demo of the algorithm was developed with the tool (Stegmayer et al., 2016) and can be accessed at: http://sinc.unl.edu.ar/web-demo/rafar/.

2.6. Performance metrics

In a standard classification task, a data input is assigned into one, and only one, of the predefined classes. There are clear definitions of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Hence, a recognizer performance can be assessed with standard metrics such as accuracy, precision, recall and F1-score (Sokolova and Lapalme, 2009). In continuous activity recognition, performance evaluation requires the comparison between a reference sequence and a recognized sequence. The activity blocks of the reference sequence and the recognized sequence may not be in a one-to-one correspondence. For example, a single block of the reference sequence can be partially detected by three shorter blocks in the recognized sequence. The definition of TP does not contemplate this kind of relation between a single input and several outputs. Thus, standard definitions of TP, FP, TN, and FN are no longer suitable. Furthermore, redefining the classification task to provide the input counts required by standard metrics can be equivocal (Ward et al., 2011). For instance, standard metrics can be computed based on frames, but they do not distinguish between serious frame errors (block insertion or deletion) and timing offsets (partially detected blocks). Also, they fail to capture common artifacts such as fragmentation, merging, and timing offsets of blocks. A simplified example of the complexity involved in the analysis of continuous activity recognition can be seen in Appendix A.

A comprehensive set of performance metrics for continuous activity recognition has been proposed by Ward et al. (2011). These metrics are based on two complementary short- and long-term timescales. They present a multidimensional and detailed description instead of a single performance number. In this way, the strengths and weaknesses of a recognizer can be assessed, avoiding ambiguity in the results. Short-term metrics are frame-based, which is the smallest fixed-length unit of time considered by the recognizer. Frame-based metrics facilitate a fine-grain analysis that resembles a continuous time analysis. In this context, a block is defined as a contiguous sequence of equally labeled frames and it has no fixed-length. Long-term metrics are block-based, which provides a different point of view, a big picture of the recognition. This is particularly valuable to detect coarse-grain artifacts and to propose modifications in the recognizer. Errors can be related to segmentation, classification, or a combination of both tasks.

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

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

3. Results

3.1. Segmentation of foraging activities

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

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

Regarding block-based metrics (right-hand side of the spider plot), these variants of the algorithm showed surprisingly large FNR and FDR. This illustrates the importance of considering long-term error metrics of the sequences, because they can reveal and characterize unseen artifacts of the recognition process. Most of the errors were due to block merging (up to 37%) and block fragmentation (up to 38%). There were also some insertions of blocks. Compared to RAFAR segmentation, the MB stage increased the block merging rate by 17% and reduced block fragmentation by 22%. The MA stage reduced fragmentation and slightly increased merging but it had no substantial effect. Block FDR was reduced but it was still high (>40%). A comparison between frame- and block-based results indicated that most blocks in the recognized sequence correctly overlapped with the blocks of the reference sequence, and most regions of non-foraging activities were not assigned to activity blocks. However, the recognized sequence was mostly

(a)

(b)

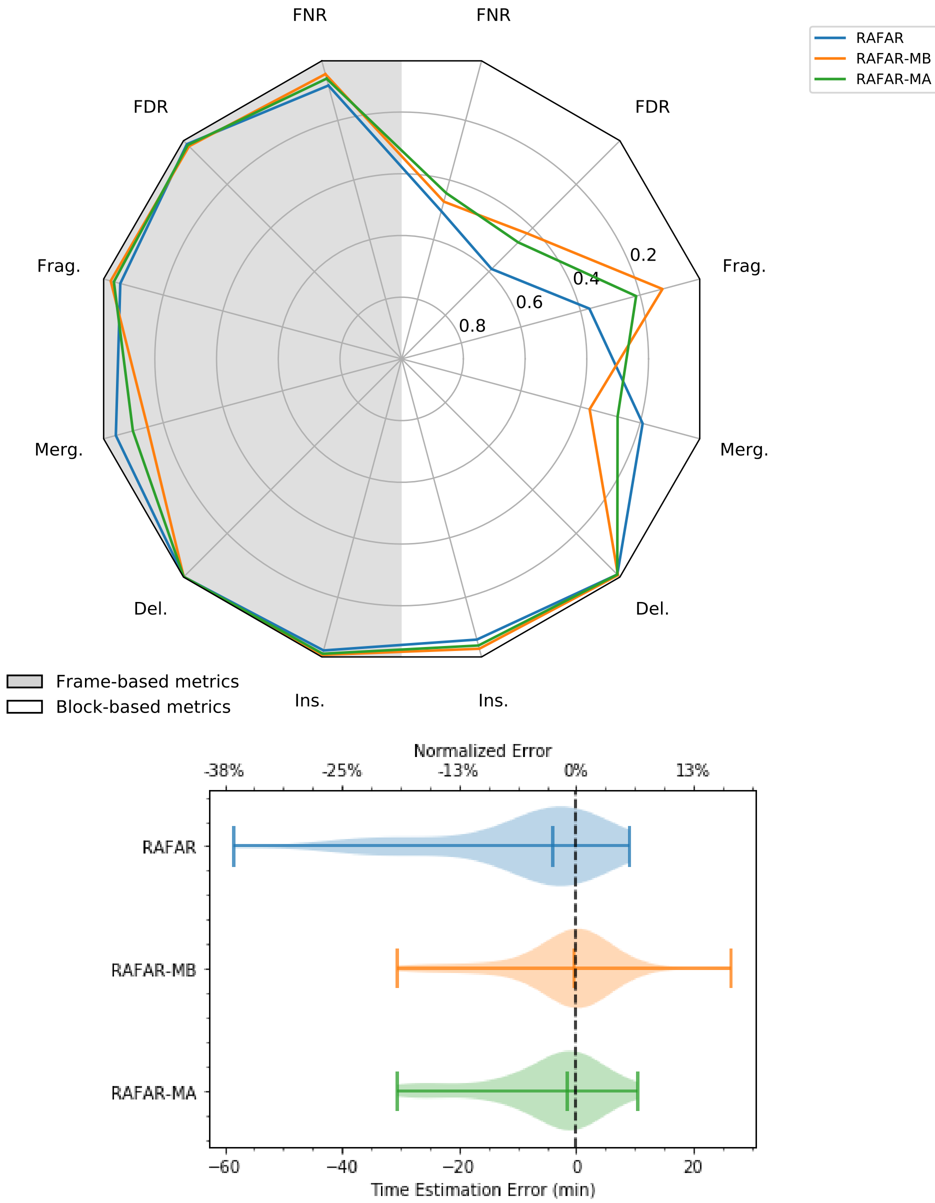


Figure 5: Segmentation of foraging activities. (a) Spider plot of frame- and block-based error metrics. The lowest errors correspond to the boundary of the polygon. (b) Violin plots of time estimation errors. The top axis was normalized with the mean duration of foraging activities (158.6 min) across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

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
S. R. Vanrell, J. O. Chelotti, J. Galli, S.A Utsumi, L. Giovanini, H. L. Rufiner & D. H. Milone; "A regularity-based algorithm for identifying grazing and rumination bouts from acoustic signals in grazing cattle"
Computers and Electronics in Agriculture, Vol. 151, pp. 392-402, 2018.

comprised of long merging blocks. This should be taken into account for the analysis of classification results and the role of a block partition stage.

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

3.2. Classification of foraging activities

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

Regarding block-based metrics, most RAFAR variants showed a FNR below 30% and FDR below 50% (Figure 6.a). Several grazing blocks of the

(a)

(b)

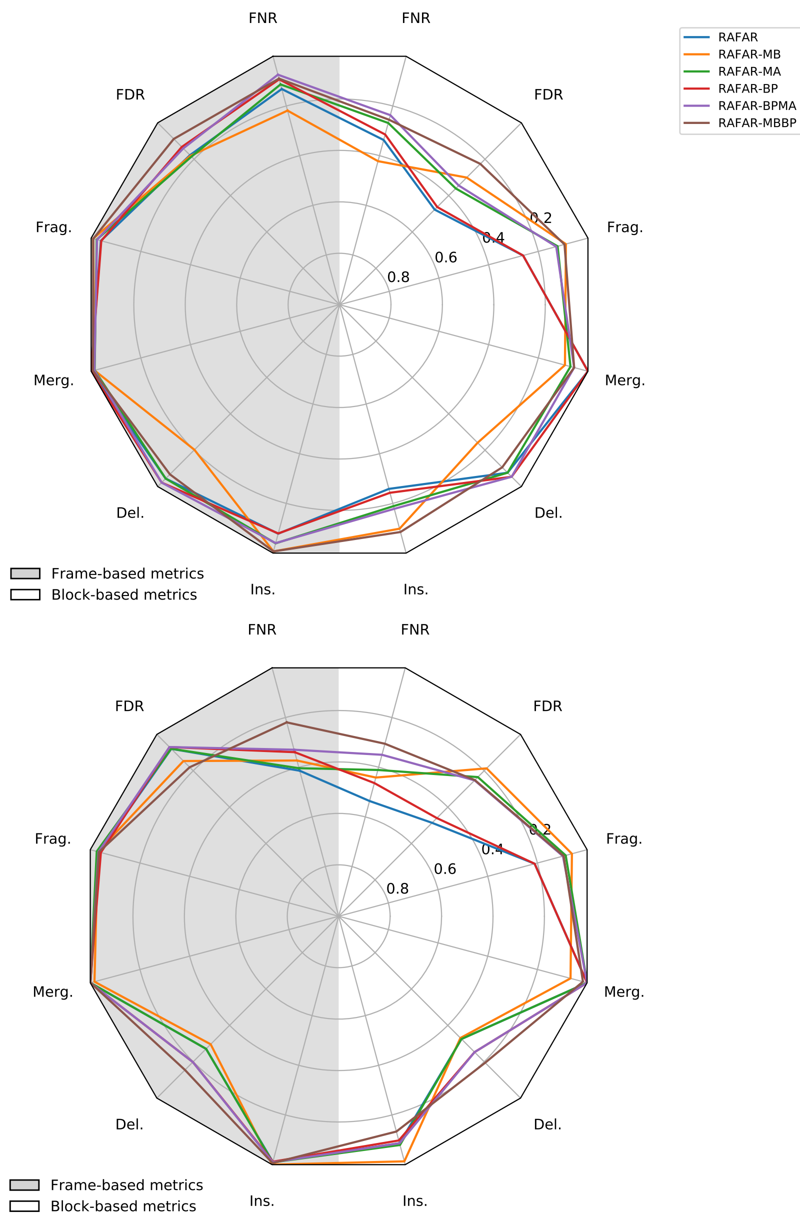
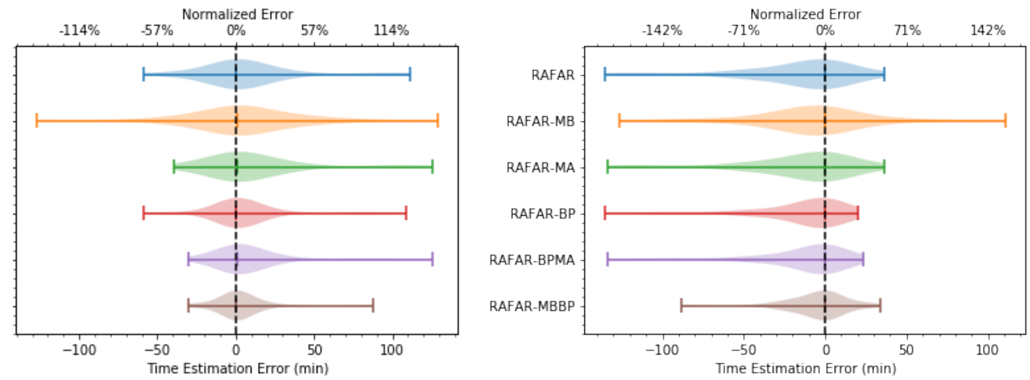


Figure 6: Spider plot of frame- and block-based error metrics for (a) grazing and (b) rumination classification. The lowest errors correspond to the boundary of the polygon. RAFAR: baseline segmentation and classification. RAFAR-MB: gap merging before classification. RAFAR-MA: gap merging after classification.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

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

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

The analysis of block results shows that two variants (RAFAR and RAFAR-BP) correctly recognized ∼50% of the rumination blocks. These variants merged no blocks and presented up to 10% of insertions. The fragmentation errors of 20% indicate that blocks in the reference sequence were partially detected. This explains the moderate FNR and FDR. The inclusion of a merging stage (RAFAR-MB or RAFAR-MA) achieved the lowest FDR and highly reduced fragmentation and insertions. RAFAR-BPMA and RAFAR-



(a) (b)

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

MBBP lowered the FNR and significantly reduced deletions. In addition, these two variants kept fragmentation, merging, and insertions at low rates. The best option was RAFAR-MBBP, which obtained a block FNR below 30% and a slightly lower FDR.

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

4. Discussion

The set of metrics proposed for analysis of the algorithm variants provided a new multidimensional view of the recognition performance at two different timescales. Frame-based metrics facilitate a fine-grain analysis that compare the reference and the recognition sequences on a timescale of seconds (1 s frame). Block-based metrics compare the sequences on a timescale of minutes or hours, which provides a big picture of the recognition. To compare and select a variant, a summary of segmentation and classification results is given in Table 2. Frame- and block-based results were summarized with the F1-score averaged across the recordings analyzed. Segmentation showed impressive frame-based scores (above 0.94), which means that the proposed algorithm was highly accurate in discriminating foraging activities from others. Block-based scores were up to 0.715. In addition to some blocks being partially detected, most activities were correctly identified. The best activity segmentation was achieved with RAFAR-MB and RAFAR-MBBP (equivalent for segmentation). Regarding classification, frame-based scores were lower than the segmentation for the same algorithm variation, which indicates that some of the correctly identified activity blocks were misclassified in the latter stage. By contrast, block-based scores obtained for classification highly improved segmentation scores. This is particularly evident for RAFAR-MBBP, which combines a merging stage that reduced fragmentation and insertions, and a block partition stage that avoided misclassifications (deletions and insertions exchange between activities). Block fragmentation and insertions were reduced because the merging stage combined the short gaps resulting from the basic segmentation, which resulted in longer blocks for classification. As expected, a single block could include a mix of grazing and rumination bouts. Thus, the partition of blocks helped to prevent the misclassification of mixed blocks. Having analyzed all the variants, RAFAR-MBBP has shown the best tradeoff and will be considered for the following comparisons.

A comparison is made of the rumination time estimation obtained by the Hi-Tag system and the proposed algorithm RAFAR-MBBP. The Hi-Tag rumination system summarizes the total time the animal spent ruminating during two-hour chunks (Schirmann et al., 2009). It provides no access to raw data on timing or duration of rumination bouts within a two-hour chunk (Goldhawk et al., 2013). Therefore, the estimations with the RAFAR-MBBP were aligned, and total duration of rumination was summarized to match the same two-hour chunks of the Hi-Tag system. The comparison was

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

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

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

made with a total of 53 two-hour chunks from all the recordings analyzed. The remaining 2-hour chunks that overlapped with times of highly-noisy environment (e.g. barn or engine noises) were discarded from this analysis.

The results of time estimation error for rumination are shown in Figure 8. The Hi-Tag system exhibited a wide distribution that ranged from heavy underestimation (-75 min) to equally high overestimation (60 min). In practical terms, these values are not negligible since they are in the same order of magnitude of the two-hour chunk analyzed. The distribution shows a tendency toward underestimation with a median of -13.55 min, in agreement with the results reported in Goldhawk et al. (2013). The histogram also shows this tendency and a uniform-like distribution between -40 and 30 min. By contrast, the proposed RAFAR-MBBP obtained a much lower error, which mostly ranged from -25 min of underestimation to 15 min of overestimation. The distribution also shows a tendency toward underestimation with a lower median of -2.56 min. The histogram shows that most errors are very close to zero and that the distribution is clearly narrower. In practical terms, this comparison suggests that the Hi-Tag system estimate rumination time with error that can be greater than 1 hour. By contrast, the proposed system provides rumination time with errors that are only a small time fraction of 1 hour.

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

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

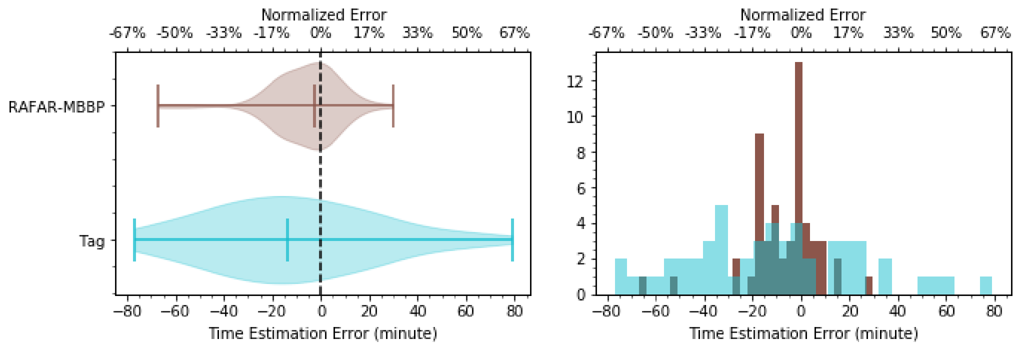


Figure 8: Time estimation error of rumination for RAFAR-MBBP (brown) and Hi-Tag (light blue). The top axis is normalized with the length of segments analyzed (2 hours). RAFAR-MBBP: gap merging before classification and partition of long blocks.

those seeked by the algorithm. However, it is expected that the distinctive regularity of interruptions during rumination will continue to allow discriminating rumination from grazing, even in patchy pastures.

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

5. Conclusions

In this study, an algorithm is proposed for segmenting and classifying foraging activity bouts in grazing cows. Remarkable results were obtained for segmentation. Frame-based F1-score was above 0.96 and the average time estimation error was below 0.5 min for signals of several hours. Classification of rumination and grazing also produced very good results. For both activities, the frame-based F1-score was above 0.89 and the average time estimation error was very close to zero. Grazing was slightly better identified than rumination. The proposed system estimated rumination duration with a much lower error than the commercial system on a free-ranging environment. The median of the error were -2.56 min for the proposed system and -13.55 min for the commercial one. The proposed system can be implemented for practical applications on similar environments to those considered in this study. Future work will focus on testing the algorithm on other types of animals and grazing environments, on developing a preprocessing stage to deal with highly noisy environments, and on applying advanced machine learning techniques to improve classification.

Acknowledgements

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

Appendix A. Example of continuous activity recognition

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

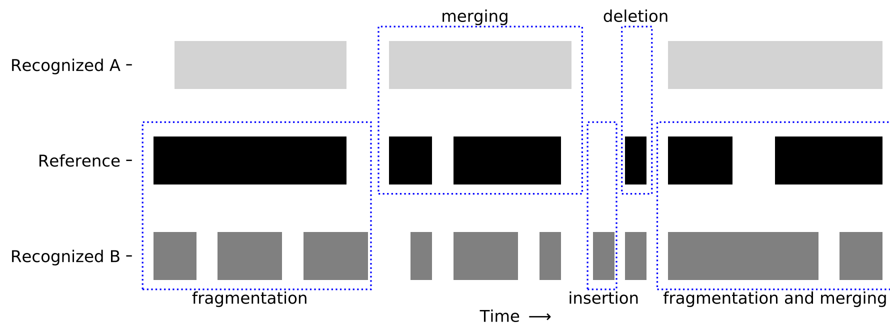


Figure A.1: Example of recognized sequences A and B compared to the reference sequence. Common artifacts found in continuous activity recognition are surrounded by dashed blue rectangles.

Appendix B. Labeling of segments

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

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

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

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

The frame- and block-based error metrics are defined in Table C.1. Frame-based metrics are defined considering the counts of true positives TP, false positives FP, false negatives FN, fragmented F, merged M, and deleted D frames in the reference sequence, and the count of inserted I frames in the recognized sequence, respectively. Frames of 1 s were considered as the smallest time unit for results analysis. Block-based metrics are defined considering the counts of total (B^{ref}), correctly detected (C), fragmented (F), merged

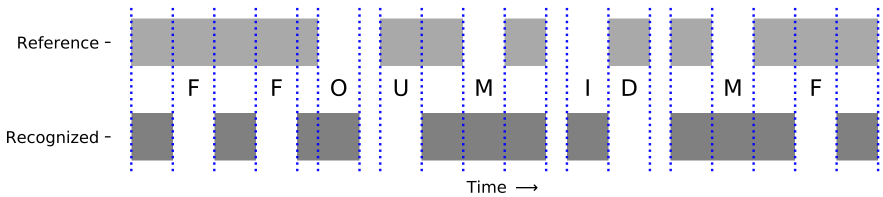


Figure B.1: Segments and boundaries defined from the comparison between the reference and the recognized sequences. Segments that would be usually classified as FN or FP are identified with the corresponding subcategory. The TP and TN labels are omitted for clarity.

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

Error metric	Frame-based	Block-based
False negative rate	$FNR_f = 1 - \frac{TP}{TP+FN} = 1 - R_f$	$FNR_b = 1 - \frac{C}{B^{ref}} = 1 - R_b$
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	$I_f = \frac{I}{TP+FN}$	$I_b = \frac{I}{B^{rec}}$

(M), and deleted (D) blocks in the reference sequence, and the counts of total (B^{rec}) and inserted (I) blocks in the recognized sequence, respectively. In addition, the standard F1-score was computed for frames $F1_f = \frac{2R_f P_f}{R_f+P_f}$ and blocks $F1_b = \frac{2R_b P_b}{R_b+P_b}$ based on the corresponding precision and recall defined in Table C.1.

References

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 (2), 439–463.

Benvenutti, M. A., Pavetti, D. R., Poppi, D. P., Gordon, I. J., Cangiano, C. A., Sep. 2016. Defoliation patterns and their implications for the management of vegetative tropical pastures to control intake and diet quality by cattle. Grass Forage Sci. 71 (3), 424–436.

Bristow, D. J., Holmes, D. S., Mar. 2007. Cortisol levels and anxiety-related behaviors in cattle. Physiol. Behav. 90 (4), 626–628.

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

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

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 (1), 96–104.

Delagarde, R., Caudal, J.-P., Peyraud, J.-L., 1999. Development of an automatic bitemeter for grazing cattle. In: Annales de zootechnie. Vol. 48. pp. 329–339.

Galli, J. R., Cangiano, C. A., Milone, D. H., Laca, E. A., 2011. Acoustic monitoring of short-term ingestive behavior and intake in grazing sheep. Livestock Science 140 (1–3), 32–41.

Galli, J. R., Cangiano, C. A., Pece, M. A., Larripa, M. J., Milone, D. H., Utsumi, S. A., Laca, E. A., 2017. Monitoring and assessment of ingestive chewing sounds for prediction of herbage intake rate in grazing cattle. Animal: An International Journal of Animal Bioscience, 1–10.

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 (6), 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., 2018. Use of sensor-determined behaviours to develop algorithms for pasture intake by individual grazing cattle. Crop Pasture Sci. 68 (12), 1091–1099.

Herskin, M. S., Munksgaard, L., Ladewig, J., Dec. 2004. Effects of acute stressors on nociception, adrenocortical responses and behavior of dairy cows. Physiol. Behav. 83 (3), 411–420.

Hodgson, J., Illius, A. W., 1998. The ecology and management of grazing systems. CAB International.

Kilgour, R. J., 2012. In pursuit of "normal": A review of the behaviour of cattle at pasture. Appl. Anim. Behav. Sci. 138 (1), 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 (1), 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 (1–2), 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 (4), 474–483.

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

Oppenheim, A. V., Schafer, R. W., Nov. 2011. Discrete-Time Signal Processing. Pearson Education.

Rutter, S., Champion, R., Penning, P., 1997. An automatic system to record foraging behaviour in free-ranging ruminants. Applied Animal Behaviour Science 54 (2-3), 185–195.

Rutter, S. M., 2000. Graze: a program to analyze recordings of the jaw movements of ruminants. Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc 32 (1), 86–92.

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 (12), 6052–6055.

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

Sokolova, M., Lapalme, G., Jul. 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45 (4), 427–437.

Stegmayer, G., Pividori, M., Milone, D. H., 2016. A very simple and fast way to access and validate algorithms in reproducible research. Briefings in Bioinformatics 17 (1), 180–183.

Trindade, J. K. D., Carvalho, P. C. d. F., Neves, F. P., Pinto, C. E., Gonda, H. L., Nadin, L. B., Correia, L. H. S., 2011. Potencial de um método acústico em quantificar as atividades de bovinos em pastejo. Pesqui. Agropecu. Bras. 46 (8), 965–968.

Ungar, E. D., Rutter, S. M., Jun. 2006. Classifying cattle jaw movements: Comparing IGER behaviour recorder and acoustic techniques. Appl. Anim. Behav. Sci. 98 (1), 11–27.

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

Welch, J. G., 1982. Rumination, particle size and passage from the rumen. J. Anim. Sci. 54 (4), 885–894.

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

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.