Automatic recognition of ingestive sounds of cattle based on hidden Markov models

Diego H. Milone[1,2,5]

Julio R. Galli[3]

Carlos A. Cangiano[4]

Hugo L. Rufiner[1,2,5]

Emilio A. Laca[6]

[1]Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral, Argentina

[2]Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina

[3]Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Argentina

[4]Estación Experimental Agropecuaria Balcarce, Instituto Nacional de Tecnología Agropecuaria, Argentina

[5]Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

[6]Department of Plant Sciences, University of California, Davis, United States

Corresponding author: D.H. Milone, e-mail address d.milone@ieee.org

Abstract

Information about ingestive events like chewing and biting is useful for estimation of intake and monitoring of grazing behaviour. We present an automatic tool to decode ingestive sounds of cattle into ingestive events. Ingestive sounds can be recorded easily and without alteration of normal grazing behaviour by placing a microphone on the forehead of the animal. However, recorded sound need to be decoded automatically for the method to be of practical use. Hidden Markov models have been successfully used to segment and classify acoustic signals. In this work we extend the use of hidden Markov models to recognize ingestive sounds of cattle. We present new findings about the spectral content of the acoustic signals and a novel language model for the recognizer. Three types of ingestive events (bites, chews and chewbites) by cows grazing tall (24.5±3.8 cm) or short (11.6±1.9 cm) alfalfa or fescue were successfully recognized. Recognition rates were 84% for tall alfalfa, 65% for short alfalfa, 85% for tall fescue and 84% for short fescue. These levels of correct classification are suitable for quantification of grazing behaviour.

1. Introduction

Accurate information about grazing behaviour is required to improve beef and milk production and to manage grazing systems sustainably. Production by grazing animals depends on intake, which is difficult to measure, particularly in time periods of hours to days. Estimation of intake based on grazing behaviour is suitable for the short-term if good estimates of bite size can be obtained. Measurements of grazing behaviour can be made by direct observation, but this can be extremely time consuming and it is very difficult to collect data over long periods of time. Thus, a number of mechanical and electronic devices have been developed to automatically record grazing behaviour (Luginbuhl et al., 1987; Delagarde et al., 1999). However, these methods are in general imprecise and invasive, because they require individual calibrations and bulky equipment on the animal.

The grazing process involves selection, apprehension, chewing and swallowing of herbage. During grazing the animal moves the jaw continuously, but two functions can be differentiated: biting, when herbage is apprehended and severed, and chewing, when herbage is comminuted to reduce particle size and increase the surface/volume ratio. Sheep and cattle can chew and bite with the same jaw movement (chewbite), which has very important implications for the correct estimations of chewing requirements of forages and routine grazing behaviour descriptions such as biting rate and chewing per unit intake.

An acoustical method to monitor ingestive behaviour of cattle (Laca and WallisDeVries, 2000) allows accurate measurement of allocation of jaw movements to elucidate the mechanisms that determine dry matter intake in the short term (minutes). Furthermore, an acceptable estimation of intake can be obtained with this method because sound energy is highly correlated with the quantity of dry matter intake (Laca and WallisDeVries, 2000; Galli et al., 2005; Galli et al., 2011). The use of this method for periods longer than a few minutes is impractical because the decoding of recorded sound by an operator is extremely time consuming. To use the acoustic method in the long term (hours), it is necessary to automate the segmentation and classification of sound signals.

Clapham et al. (2011) recently proposed a decoding approach based on high-frequency narrow-band filtering and thresholding applied to the raw (time-domain) acoustic signal. They reported high rates of detection for bites, but the system required manual calibration for each animal and forage. We propose hidden Markov models (HMM) as a framework to explicitly model and classify ingestive sounds of cattle, given the robust results achieved in similar tasks by Reby et al. (2006), Trifa et al. (2008) and Milone et al. (2010). Hidden Markov models were tested to segment and recognize mastication sounds in sheep (Milone et al., 2009) and preliminary results were reported for cattle (Milone et al., 2008).

In this work we extend the model for its application to cattle and provide a new language model to improve the recognition performance. The main hypothesis is that the sound contains all the information needed to correctly classify the three most important sounds of chewing cattle (i.e. bites, chews and chewbites). The objectives of this work are to explore signal processing methods to extract the main characteristics of ingestive sounds of cattle, to provide statistical models to classify this sound based on the spectral dynamics and to develop statistical models for the sequences of chewing events.

This article is organized as follows: Section 2 details the acquisition of ingestive signals and shows a preliminary comparison between different events in the frequency domain. Section 3 presents the feature extraction and the classifier design. Results are presented and discussed in Section 4 and finally the conclusions are given in Section 5.

2. Materials

Fieldwork was performed at the Campo Experimental J. F. Villarino, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Argentina. The project was evaluated and approved by the Committee on Ethical Use of Animals for Research of the Universidad Nacional de Rosario. Sound signals from dairy cows grazing alfalfa or fescue of two different heights (tall, 24.5±3.8 cm or short, 11.6±1.9 cm) were recorded in individual grazing sessions during a period of 5 days from 17 to 21 February 2004. Forage species were selected because they differ greatly in structure and neutral detergent fibre content (alfalfa, 360±11 g/kg and fescue, 631±6 g/kg), which are factors that have strong influence on the sound of chewing (Duizer, 2001). Two 4-6 year-old lactating Holstein cows weighing 608±24.9 kg, previously tamed and trained were used. Two wireless microphones (Nady 151 VR, Nady Systems, Oakland, CA, USA) were randomly assigned to animals each day. The microphone was placed facing inward on the cow's forehead protected by foam rubber (Milone et al., 2009). The distance between the wireless microphone and the receiver was 2-3 meters.

Each cow grazed plants in pots that were firmly attached to a board placed inside a barn. Behaviour was recorded with an analog video camcorder (Sony CCD-TR517), and then coded in MPG format at 25 frames per second. The sound from the wireless microphone was recorded on the tape soundtrack (16 bits, 44.1 kHz). A standard beeping sound was produced every 10 s to equalize sound intensity across recordings. The beeping sound was later filtered with a notch filter at 4100 Hz (Kuc, 1988). In spite of the fact that the recordings were obtained indoors they contain various types of environmental noises, such as birdsong. However, denoising was not applied; that is, signals were fed to the recognizer as originally recorded. A total of 50 signals were obtained: 15 from tall alfalfa, 11 from short alfalfa, 12 from tall fescue and 12 from short fescue. On average, for each pasture/height the signals contained approximately 13 minutes of recording and around 800 events (13% bites, 64% chews and 23% chewbites).

Experts in ruminant grazing behaviour, well trained in recognition of ingestive sounds, viewed and listened to the recordings and identified the events with labels on the plot of the sound wave (Figure 1).

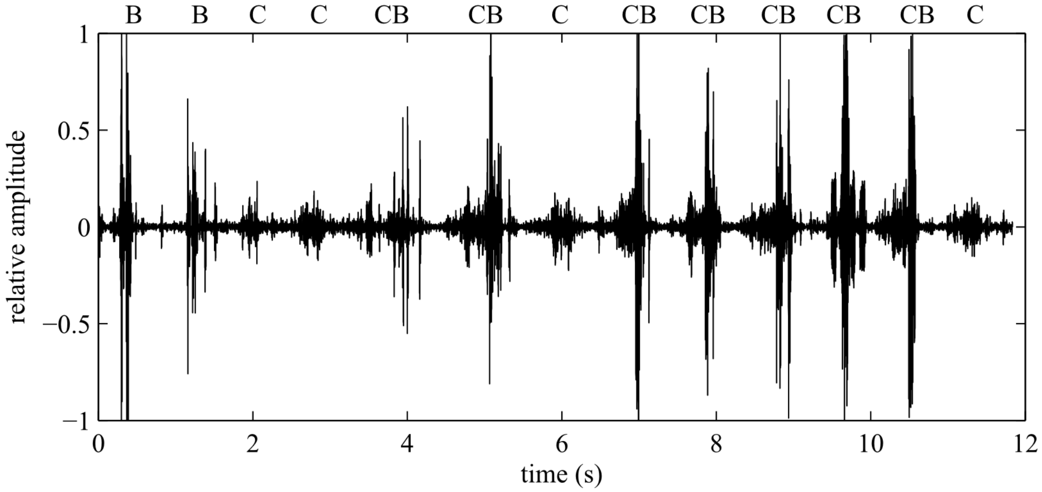

Figure 1. Fragment of an acoustic signal taken from tall alfalfa with the corresponding sequence of hand-labelled events.

3. Methods

Signals analyzed consist of sequences of events -bites, chews and compound chewbites- separated by “silences”. Every event has a duration between 200 and 500 ms. These signals can be considered approximately stationary by segments between 20 and 80 ms. That is, for such small segments, we consider that the statistical properties of the acoustic signal remain unmodified. Thus, it is possible to use some analysis tools (like the discrete Fourier transform) that are applicable to stationary signals only. These fixed-length segments defined for analysis are commonly referred to as “frames” in the jargon of signal processing.

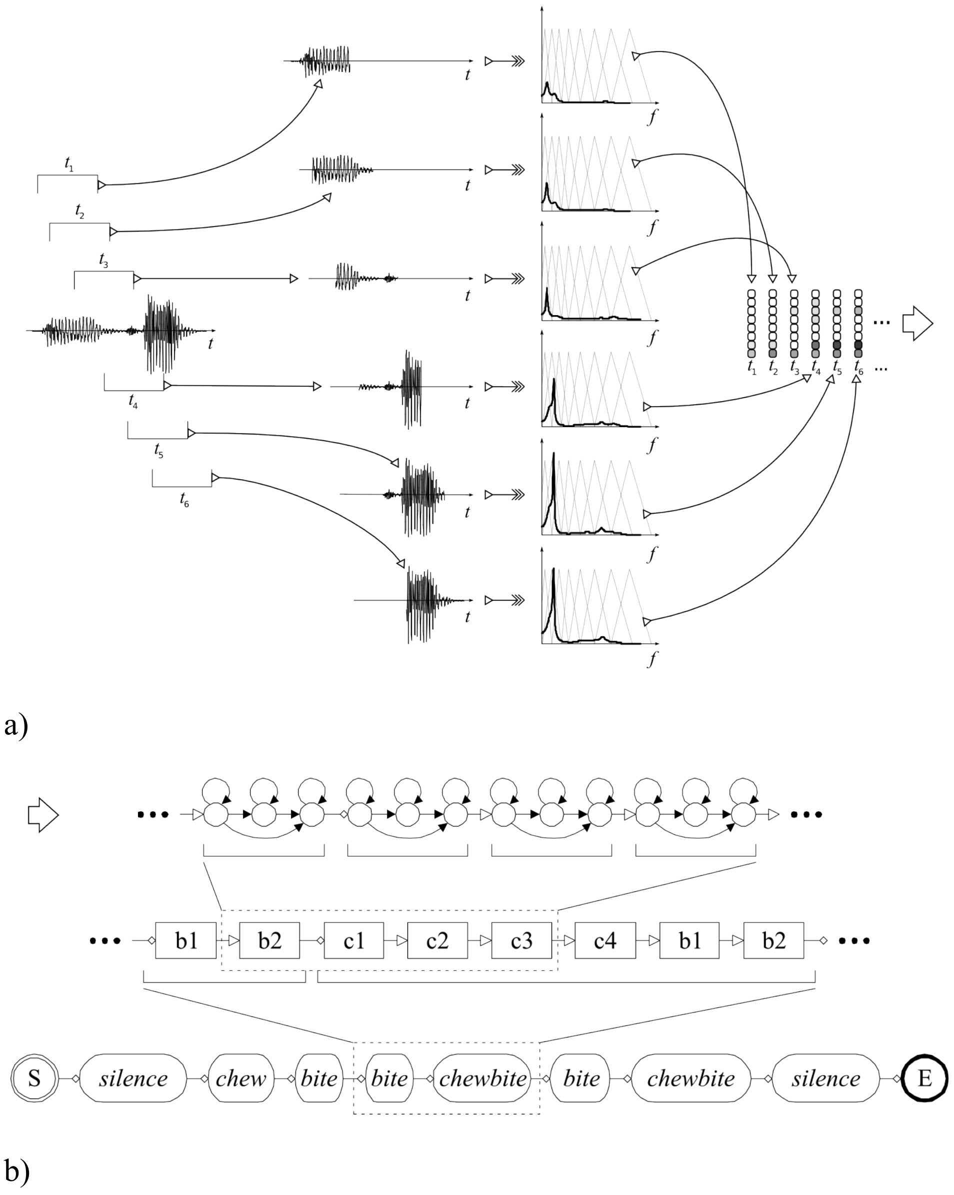

Figure 2. The recognition process: a) feature extraction; b) acoustic models (states of the hidden Markov models), ingestive sub-events (b1-2 are sub-events of the bite model and c1-4 are sub-events of the chew model) and language modelling for sequences of events.

The recognition system can be separated into three main parts: feature extraction (Figure 2.a), statistical acoustic model (top of Figure 2.b) and statistical language model (middle and bottom of Figure 2.b). In the feature extraction part, the original signal is translated into a time series of vectors describing the spectral composition of windows of the signal. This step was performed using standard techniques (Rabiner and Juang, 1993) for which we selected the Hamming window with steps and durations leading to the highest recognition rates within reasonable ranges. Length of vectors, which is determined by the characteristics of filter banks used in this feature extraction step, was selected using a similar procedure of looking for the lengths that led to high recognition rates (Young et al., 2005).

In the second part, the acoustical model is built using hidden Markov models. These models are statistical methods to optimally estimate the sequence of states of a system that is not directly observable (thus the term “hidden”), but that has observable characteristics whose statistical distributions depend on the state (Huang et al., 2001). In our case and simplifying a bit for the sake of clarity, the states are the components of ingestive events and the observable characteristics are their acoustic spectra. Hidden Markov models assume that the current state depends only on the previous one, regardless of how the system arrived at the previous state (Rabiner and Juang, 1993). An HMM consists of a graph of states and transitions, a matrix of parameters describing the transition probabilities between any two states, and a set of parametric functions describing the statistical distribution of observable characteristics for each state. The unobservable sequence of states is estimated by maximizing the probability of obtaining the observed sequence of vectors. Because we do not know the actual values of the parameters, they have to be estimated using “training” data sets for which the sequences of true events were observed. In this work, we consider that the decoding of the video and audio by experienced observers yields the true events. The Baum-Welch algorithm was used for parameter estimation (Rabiner, 1989). We considered that this iterative estimation process converged when the average log-likelihood of the parameters of the model given the data do not change much from one iteration to the next. All the recognition tests were implemented using the HTK toolkit[1].

The third part of the recogniser is the long-term statistical model or “language” model (LM). The general structure of the proposed system resembles that of a speech recognition system, where

[1] http://htk.eng.cam.ac.uk

phoneme models are replaced by ingestive sub-events and word models by complete events. As in the speech case, the LM captures the long-term dependencies and constrains the possible sequences (bottom of Figure 2.b). The modelled ingestive events, like words, were: *bite, chew* and *chewbite*. In addition, a model for silences was introduced between ingestive events. Sub-events, like phonemes, were defined within each event model. For example, the definition of the bite event could be *bite* = [b1, b2], or *bite* = [b1, b2, b3], where bi are sub-events of the bite event. Each sub-event consisted of several states, with transitions probabilities and observations distributions.

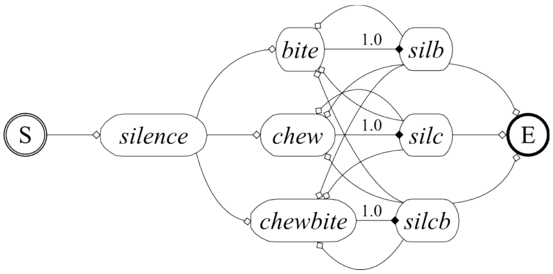

Figure 3. Diagram of the language model proposed for the recognition of ingestive sounds of cattle.

At the beginning of the model development, a simple bi-gram was used as language model (Milone et al., 2009). In this case, the probabilities of all the possible connections between the events were estimated. However, in this language model it was noticed that the *silence* event follows every one of the other events. So this bi-gram resulted in high probability between each event and *silence*, and did not allow links between the *bite*, *chew* and *chewbite* events. As a solution to this limitation, a more complex model was designed. The proposed bi-gram for the LM (Figure 3) included three types of additional silences where each silence is associated only with the corresponding event: *silb* is always after *bite*, *silc* is always after *chew* and *silcb* always after *chewbite*.

The process above results in a series of models for the different forage treatments. The final performance of the complete system set up with the best configuration was analysed using two recognition measures. Both measures compared the true sequence of events with the recognizer output. The two sequences of labels were aligned and then performance measures were calculated. For example, consider the following aligned sequences:

Reference seq.:	bite	chew	chew		chew	chew	bite	chew	chewbite
Recognised seq.:	bite	chew	chew	bite	chew	chew	chew	chew	________

The second bite of the recognised sequence is an insertion and the fifth chew is a substitution. The last event from the reference sequence has not been recognised so it is a deletion. Based in these counts, the first measure was the percentage of events correctly recognized $R = (N-(D+S))/N$, where D is the number of omitted events, S is the number of substituted events and N the total number of events in the reference transcription provided by the experts[2]. The second measurement, called accuracy, takes into account the insertions and is defined as $P = (N-(D+S+I))/N$, where I is the number of insertions. Note that in this way the sequences of labels, and not their respective timings, were aligned.

To evaluate the generality of the model, a hold-out cross-validation method was used (Duda et al., 1999). For each validation a complete grazing session was left out for testing while the rest were used for training. Performance measures (R and P) were computed for all the partitions and then the average over all partitions was computed to obtain the global recognition rates. The number of cross-validation partitions was equal to the number of recorded signals for each species. For example, tall alfalfa model was trained with the signals 2, 3, ..., 15 and tested with the signal 1. Then, a new model was trained with the signals 1, 3, ..., 15 and tested with signal 2. When all the signals were used in the test, the final result for the tall alfalfa was computed as the average recognition rate over the 15 tests.

[2] Two experts decoded records of signals. Detections agreed in 100% for bites, 98.2% for chews and 99.1% for chewbites. There were confusions of 0.9% of chewbites by bites and 1.8% of chew by chewbites. In addition, there were 2.7% of insertions, 0.9% of deletions. Thus, the total human accuracy in this case was 93.6%.

4. Results and discussion

The main differences among average spectra of 30 instances of each event are between 50 and 250 Hz (Figure 4). Bites and chewbites are very similar between 250 and 500 Hz, but both have more energy than the chew events in this range of frequencies. This qualitative analysis suggests that the main information to discriminate the events is below 500 Hz.

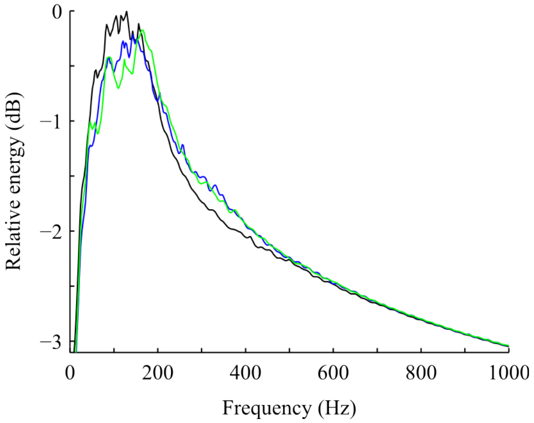

Figure 4. Average spectrum from 30 samples of each class of ingestive events: bites in blue, chews in black and chewbites in green.

Clapham et al. (2011) eliminated all frequency components below 600 Hz in order to remove wind noise prior to any analysis. We reduced the impact of the wind and other environmental noises by placing the microphone facing and pressed against the animal's forehead with a rubber foam cover. The fact that we found event-discriminating characteristics at very low frequencies, whereas Clapham et al., (2011) found them at the higher end of the spectrum suggest that a combination method including both extremes of the spectrum would be even more successful.

Highest correct recognition rates were obtained with models and parameter values that differed among forage treatments. After testing maximum frequencies between 500 and 3000 Hz, with 10 to 22 filters in each tested range, best results were obtained with 10 filters in the range from 0 to 500 Hz, except for short alfalfa, which required 22 filters in the range from 0 to 700 Hz. For window size and step, options between 20 and 80 ms were tested, and highest rates of correct recognition were obtained with a window step of 40 ms and a window size of 60 ms in all species and heights. It was also determined that four states yielded the highest recognition rate after testing models with 3-8 states in each HMM. Finally, it was found that the optimal for tall alfalfa was two sub-events in *bite*, four in *chew*, and three in *chewbite*. In the case of short alfalfa the best model contained four sub-events in *bite*, four in *chew* and five in *chewbite*. For all tests with fescue (tall and short), the optimal configuration was achieved using three sub-events in *bite*, four in *chew* and four in *chewbite*.

The models show good generality, except that recognition rates were quite low for short alfalfa (Table 1). Cows grazing alfalfa produced less “crunchy” sound than with other forages (Rutter et al., 2002). Therefore, recordings for short alfalfa have a lower signal-to-noise ratio. This method clearly distinguished among types of jaw movement: bites, chews or chewbites (Table 2). Overall, recognition was good for all forages except short alfalfa, which had a particularly high number of chewbites misclassified. The most frequently confounded events were chews and chewbites in tall pastures, chewbites and bites in short alfalfa and chewbites and chews in short fescue. The confusion of chewbites can be ameliorated by incorporating a measure of event duration, given that chewbites typically produce longer sounds because they include a sequence of a chew and one bite.

Species	***Height***	***R%***	***P%***
Alfalfa	Tall	84	79
	Short	65	57
Fescue	Tall	85	82
	Short	84	79

Table 1. Average recognition percentages obtained by cross-validation. See text for definition of R and P.

	%	Bite	Chew	Chewbite
Tall alfalfa	Bite	**79**	11	09
	Chew	03	**88**	09
	Chewbite	02	03	**94**
Short alfalfa	Bite	**76**	16	08
	Chew	05	**90**	05
	Chewbite	23	16	**61**
Tall fescue	Bite	**83**	00	17
	Chew	01	**93**	07
	Chewbite	01	04	**94**
Short fescue	Bite	**90**	09	01
	Chew	00	**99**	01
	Chewbite	02	07	**91**

Table 2. Discrimination of bites, chews and chewbites in the different pastures. Each row contains the distribution of true events over the categories into which they were classified by the recognizer. For example, in tall alfalfa 11% of the bites were incorrectly "recognised" as chews.

Clapham et al. (2011) reported detection of bites that was 95% correct. Their quantitative results are not directly comparable to ours because they focused exclusively on detection of bites, whereas we detected and classified all jaw movements. The two studies also differed in type and height of pasture, number of events analysed, duration of records and validation method. The recognition method proposed by Clapham et al. (2011) requires periodic calibration for individual animals, whereas the method we present uses one calibration for all animals.

Correct discrimination of chewbites (Laca and WallisDeVries, 2000) is important to correctly quantify the amount of chewing required by different forages (Galli et al, 2005). Chewing is physiologically important not only to promote the ruminal fermentation of fibre, but also to maintain rumen pH and supply of N recycled through saliva (Sauvant, 2000). Thus, cattle have a minimum chewing requirement to maintain health. Grazing cattle may allocate a high proportion of their jaw movements to chewbites, which are visually indistinguishable from bites because herbage already in the mouth is chewed while fresh herbage is severed in the same single jaw movement. Chewbites have been observed in giraffes (Ginnett and Demment, 1995) and sheep (Galli et al., 2010). For this reason, the acoustic method is considered to be more reliable and accurate for counting “functional” bites (bites + chewbites) and chews (chews + chewbites) than other methods (Ungar et al., 2006).

More information is required for reliable extrapolation of these results to cattle differing in breed, age and size. Chewing activity of cattle presents pronounced individual differences by physiological state, age and body size (De Boever et al., 1990), but these factors do not necessarily cause differences in chewing sound. Extrapolation of this method is facilitated by its low computational/time cost (less than 1 time the length of the signal, in a standard personal computer). Moreover, the explicit modelling with HMMs offers the possibility of automatic classification of behaviours beyond ingestive events, including rumination and drinking.

5. Conclusions

The HMM-based recognition system was able to automatically segment and classify ingestive sounds of grazing cattle. Models were tuned for optimal performance using a compound model with different levels of analysis, from the acoustics of sub-events to the long-term dependence given by the intake language model. Recognition rates up to 85% were obtained with two different pastures and heights. Moreover, the recognizer was able to perform well in cross-validation, which denotes robustness in the training algorithms and proposed models. Automatic acoustic monitoring of ingestive behaviour is valuable to assess animal health in a manner that cannot be achieved with other methods, not even by direct visual observation, because chewbites include chews and are not visually different from bites. In order to maintain rumen health, ruminants have a minimum daily chewing requirement. This study provides a basis for future work on the complete automation of recording, segmentation and classification of ingestive sounds of cattle intake, for behaviour studies and a wider application of the acoustic method.

Acknowledgements

The authors would like to thank Marcelo Larripa and Mariela Pece for their assistance in animal management and Soledad Padron for adapting part of the recogniser scripts. This work has been funded by ANPCyT and UNL, under Projects PICTO-UNL 36214, PAE-PICT 52 and CAI+D R4-N14, and UNR under Project AGR113.

References

Clapham, W.M., Fedders, J.M., Beeman, K., Neel, J.P.S.: Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle, Computers and Electronics in Agriculture, 76 (2011) 96-104.

De Boever, J.L., Andries, J.I., Brabander, D.L.D., Cottyn, B.G., Buysse, F.X.: Chewing activity of ruminants as a measure of physical structure - a review of factors affecting it. Animal Feed Science and Technology 27 (1990) 281-291

Delagarde, R., Cudal, J., Peyraud, J.: Development of an automatic bitemeter for grazing cattle. Ann Zootech (1999) 329-339

Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. 2 edn. John Wiley and Sons (1999)

Duizer, L.: A review of acoustic research for studying the sensory perception of crisp, crunchy and crackly textures. Trends in Food Science and Thechnology, 12 (2001) 17-24

Galli, J., Cangiano, C., Laca, E., Demment, M.W.: The sound of chewing. In XX International Grassland Congress, The Netherlands, Wageningen Academic Publishers (2005) 490

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

Ginnett, T.F., Demment, M.W. 1995. The functional response of herbivores: analysis and test of a simple mechanistic model. Funct. Ecol. 9, (1995) 376–384.

Huang, X., Acero, A., Hon, H.: Spoken Language Processing: a guide to theory, algorithm and system development. Prentice Hall, USA (2001)

Kuc, R.: Introduction to digital signal processing. McGraw-Hill Book Company (1988)

Laca, E., WallisDeVries, M.F.: Acoustic measurement of intake and grazing behavior of cattle. Grass Forage Sci. 55 (2000) 97-104

Luginbuhl, J., Pond, K., Burns, J.: A simple electronic device and computer interface system for monitoring chewing behavior of stall-fed ruminant animals. Journal Dairy Science 70 (1987) 1307-1312

Milone D.H., Padrón, M.S., Galli J., Cangiano C., Rufiner H.L.: Automatic recognition of ingestive sounds of cattle based on hidden Markov models. In XXXIV Conferencia Latinoamericana de Informática, Argentina, SADIO, (2008) 1130-1138.

Milone D.H., Rufiner H.L., Galli J., Laca E., Cangiano C.: Computational Method for Segmentation and Classification of Ingestive Sounds in Sheep. Computers and Electronics in Agriculture, 65(2) (2009) 228-237.

Milone D.H., Di Persia, L.E., Torres, M.E.: Denoising and Recognition using Hidden Markov Models with Observation Distributions Modeled by Hidden Markov Trees, Pattern Recognition, 43,

(2010) 1577-1589.

Rabiner, L., Juang, B.: Fundamentals of Speech Recognition. Prentice Hall (1993)

Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. Proc. IEEE 77(2) (1989) 257-286

Reby, D., André-Obrecht, R., Galinier, A., Farinas, J., Cargnelutti, B.: Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (Cervus elaphus) stags, The Journal of the Acoustical Society of America, 120(6), (2006) 4080-4089

Rutter, M.S., Ungar, E.D., Molle, G., Decandia, M.: Bites and chews in sheep: Acoustic versus automatic recording. Xth European Intake Workshop-Techniques for investigating intake and ingestive behaviour by farmed animals. Reykjavic, Iceland Agricultural Research Institute, (2002) 22-24

Sauvant, D. Granulométrie des rations et nutrition du ruminant. INRA Prod. Anim. 13 (2): (2000) 99-108.

Trifa, V.M., Kirschel, A.N.G., Taylor, C.E., Vallejo, E.E.: Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. The Journal of the Acoustical Society of America, 123(4), (2008) 2424-2431

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

Young, J., Evermann, G., Gales, M., Hain, T., Kershaw, D., Moore, G.: The HTK Book. (2005)