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

Diego H. Milone[1,5,2], María Soledad Padrón[1], Julio R. Galli[3], Carlos A. Cangiano[4], and Hugo L. Rufiner[1,5,2]

[1] Facultad de Ingeniería y Cs. Hídricas, Univ. Nacional del Litoral, Argentina
[2] Facultad de Ingeniería, Univ. Nacional de Entre Ríos, Argentina
[3] Facultad de Ciencias Agrarias, Univ. Nacional de Rosario, Argentina
[4] EEA Balcarce, Instituto Nacional de Tecnología Agropecuaria, Argentina
[5] Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

Abstract. An automatic tool capable of informing when bites and chews occurred, using only the acoustic signal, is presented. This information is useful for the estimation of intake and the monitoring of grazing behaviour. Additionally, the sound recording is easy to implement through the placement of a wireless microphone in the forehead of the animal, without altering its normal behaviour. Hidden Markov models have been used to segment and classify acoustic signals with good results. In this work we propose the use of hidden Markov models to recognize masticatory sounds of cattle. Three types of masticatory events from cows grazing pastures of two plant species (alfalfa and fescue) and two heights (tall and short) were successfully recognized. Recognition rates were 83.95% for tall alfalfa, 65.33% for short alfalfa, 84.68% for tall fescue and 83.68% for short fescue.

1 Introduction

To improve beef and milk production from ruminants (cattle and sheep) accurate information about grazing behaviour is required. The level of production obtained from grazing animals depends on their ability to ingest an adequate diet to meet their nutrient requirements for maintenance, growth, production and reproduction. 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. So a number of mechanical and electronic devices have been developed to automatically measure the grazing behaviour [1, 2]. However, these methods are in general imprecise and invasive.

The grazing process involves not only the ingestion of forage, but also the selection, apprehension, chewing and swallowing. During the ingestive process the animal moves the jaw continuously, but two phases can be differentiated: the biting, when grass enters the animal mouth and the chewing, when it is comminuted to reduce in size, increasing the surface/volume ratio. Sheep and cattle can also chew and bite with the same jaw movement (chew-bite), which has very important implications for the estimations of chewing requirements of forages.

At the end of the 80s, an acoustical method was proposed for monitoring ingestive bites and chews of cattle [3]. This method permits accurate measurement of allocation of jaw movements to elucidate the mechanisms that determine dry matter intake in the short term (minutes). Furthermore, acceptable estimation of intake can be obtained, given that the energy of the sound is highly related to the quantity of dry matter intake [4–6]. To use the acoustic method in the long term (hours) is necessary to automate the recognition of sound signals.

Sounds recorded during the grazing process contain a sequence of events interspersed by silences, where each event is the result of one of the previuosly mentioned stages. In this work the hidden Markov models (HMM) are proposed as a framework for mastication sound modelling because of the good results achieved in similar problems. Recently, HMM were tested in mastication sounds in sheep [7, 8]. In this work we have extended the model for its application to cattle and provided a new language model to improve the recognizing performance.

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

2 Materials

Sound signals were captured by experts in ingestive behaviour from dairy cows grazing alfalfa and fescue with two different heights (tall and short). The signals were recorded in various individual grazing sessions during five consecutive days with two microphones randomly assigned every day.

The recordings consists of placing the microphone on the cow's forehead (facing inward) and letting the animal feeds the plants from pots fixed to a board. This sound recording was done simultaneously with video recording that captures the animal movements. A beep sound was reproduced every 10 s to be able to coordinate both capture processes. The beep sound was filtered then with a notch filter to eliminate the 4100 Hz component [9].

Two lactating Holstein cows of 4-6 years old, weighing 608±24.9 kg, previously tamed and trained were used. The total number of signals was 50: 15 from tall alfalfa, 11 from short alfalfa, 12 from tall fescue and 12 from short fescue. On average, each pasture/height contained approximately 13 minutes of recording and around 800 events. The two species of forages were selected because one of them is a grass (fescue) and the other a legume (alfalfa), allowing to compare different textures and structures. The texture has strong influence in the sound of chewing [10].

The experts labeled the signals identifying the sequence of events, from the analysis of the sound and the video at the same time (Figure 1). The signals

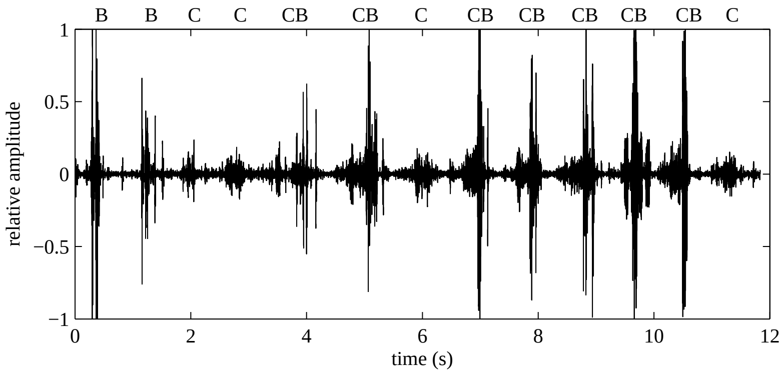

Fig. 1. Example of a sequence of labeled events taken from tall alfalfa.

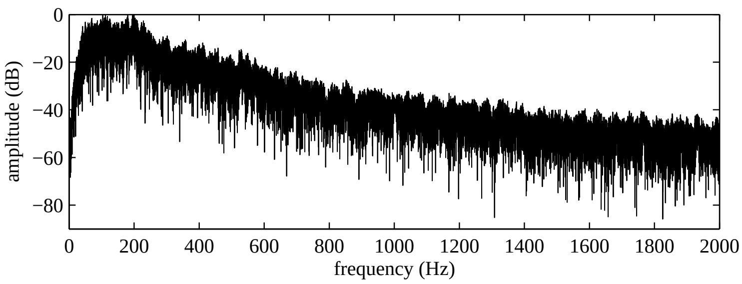

Fig. 2. Example of the spectrum for masticatory sounds.

consist of a sequence of events –chew, bites and compound chew-bites– separated by silences (very low energy events). Every event has a duration between 200 and 500 ms. These signals can be considered stationary by segments of approximatelly 20 ms.

Fourier Transform was applied to make the frequency analysis of the signals (Figure 2). There was not significant amplitudes for frequencies over 500 Hz, determining the range of interest for the design of the experiments.

Figure 3 shows a preliminary comparison between bites, chews, and chew-bites events in frequency domain, showing that the peaks occurred at different frequencies for each event.

3 Methods

The principal objective of the recognition methods is to clarify the complex mechanisms in the process of making decisions and automatize these functions using computers [11]. In this work, recognition consists of the segmentation and classification of masticatory events. The recognition system can be separated into two main parts: the feature extraction and the dynamical classifier.

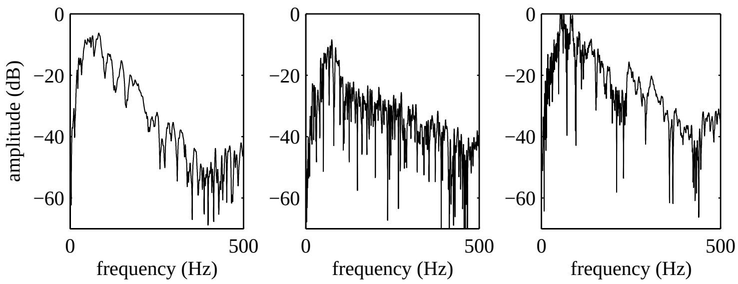

Fig. 3. Sounds of events in frequency domain. From left to right: bite, chew and chew-bite.

In a first stage the frames of the acoustic signals were extracted and transformed into a sequence of feature vectors. These vectors have a considerably smaller size than the original frames, but they can still describe the more relevant information.

Short term analysis methods allow to use the joint time-frequency information. For this analysis the Hamming window was used [9].The step of analysis is the time between the beginning of two successive windows, and in general the step is smaller than the size of the windows.

For the parameterization a filter bank was used, a collection of filters spanned along the frequency domain. In this case the filters were triangular and based in the Fourier transform designed to obtain a resolution similar to the mel scale [12]. Band-limiting was used to reject unwanted frequencies. Additionally to the coefficients that resulted of applying the filter bank to the windowed signal, other coefficients were added to provide more information from the original signal: the *normalized energy* and *delta coefficients*, the first derivative respect to time of the basic static parameters.

The HMM are statistical methods suitable for characterizing observable data by estimating an appropriated parametric model. HMM are finite state automata based on the hypothesis that the data is generated by a random process with temporal dependencies of first order [13].

A Markov chain is a sequence of states where each state is in deterministic relation with an observable event. HMM are extensions of Markov chains where states are not directly visible. The observation is a probabilistic function of the state, so each state has a probabilistic function over the possible outputs [14].

If the observations do not come from a finite set, but from a continuous one, continuous HMM can be defined. In semicontinuous-HMM (used in this work), the observation probabilities share a Gaussian set and for each state different weights are assigned in the mixture. To model temporal series, left to right HMM are used, where no backward transitions are permitted (Figure 4). Given the observation vectors, the model parameters must be estimated from data. The Baum-Welch algorithm was used for this training stage [15].

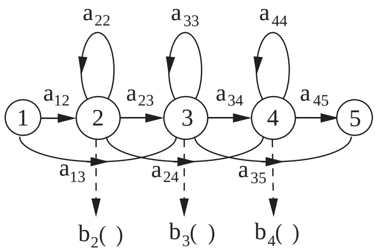

Fig. 4. Left to right hidden Markov model. a_{ij}: transitions probabilities; $b_j(\cdot)$: observation distributions.

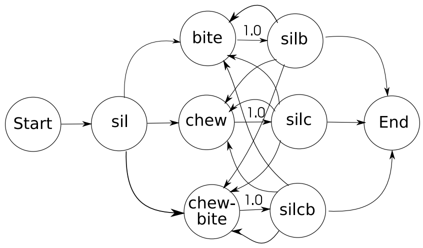

Fig. 5. Diagram of the language model proposed for the recognition of cattle masticatory sounds.

The general structure of the proposed system resembles to that of a speech recognition system, where phoneme models are replaced by masticatory subevents and word models by complete events. As in the speech case, the language model (LM) captures the long-term dependencies and constrains the possible sequences. The modeled masticatory events were: *chew*, *bite* and *chew-bite*. Silence event (*sil*) was modeled for separating these masticatory events. Moreover, sub-events were modeled. As an example, the definition of the *bite* event is $bite = [b1, b2]$ where *b1* and *b2* are sub-events of the *bite* event. Each sub-event consisted of several states, where transitions probabilities and observations distributions were estimated.

At the beginning of the model development, a simple bigram was used as language model [14]. 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 *sil* event follows every one of the other events. So this bigram resulted with high probability between each event and the *sil*, but did not allow links between the *bite*, *chew* and *chew-bite* events. As a solution to this limitation, a more complex model was designed, including three types of additional silences, *silb*, *silc* and *silcb*, where for example *silb* was a silence after a *bite* event, and so on (Figure 5).

Table 1. Overall averages of the train/test partitions.

Species and height	$R\%$	$P\%$
Tall alfalfa	83.95%	78.69%
Short alfalfa	65.33%	56.66%
Tall fescue	84.68%	82.38%
Short fescue	83.68%	78.87%

4 Results and discussion

The experimental methodology consisted of two main steps. In the first step, several experiments were performed to define the general structure of the recognizer. In the second step the models were trained and the recognition rates were computed with the testing data.
To define the general structure of the recognizer we considered:

- Alternatives in the initialization of the models
- The previous mentioned methods for feature extraction, including
  - Size and step of the window used
  - Number of filters
  - Low and high band-limiting frequencies
- Number of states for each sub-event
- Number of sub-events for each type of event
- Language model alternatives

The efficiency of the recognizer was analyzed using two performance measurements. Both compare the labels generated by the expert with the recognizer output. First, the two sequences of labels are aligned with a dynamic programming method and then the measures were calculated. The first measure was the percentage of labels correctly recognized $R\% = \frac{N-(D+S)}{N} \times 100\%$, where D is the number of omitted events, S is the number of replaced events and N the total number of labels in the transcription file provided by the expert. The second measurement takes into account the insertions and is defined as $P\% = \frac{N-(D+S+I)}{N} \times 100\%$, being I the number of insertions.

To evaluate the generalization ability of the obtained model, a cross-validation method was used [16]. The way for partitioning the data is related to the way the sound has been recorded. Taking into account that each recording session correspond to one audio file, for each validation test a complete session was left apart and the others 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 validation results with the proposed method for each specie and height are shown in Table 1. The number of cross-validation partitions was dependent on the number of recorded sessions for each species: 13 for tall alfalfa, 9 for short alfalfa, 10 for tall fescue and 10 for short fescue.

Table 2. Discrimination of bite, chew and chew-bites in the different pastures. Columns: recognized events; Rows: reference events.

	bite	chew	chew-bite
Tall alfalfa			
bite	77	11	9
chew	10	272	28
chew-bite	6	8	229
Short alfalfa			
bite	81	17	8
chew	9	170	9
chew-bite	19	13	50
Tall fescue			
bite	57	0	12
chew	3	327	23
chew-bite	2	7	151
Short fescue			
bite	62	6	1
chew	1	329	3
chew-bite	4	12	153

Good results were obtained for the two species of forages. The models show an important generalization capability, although recognition rates were quite low for short alfalfa. This legume (alfalfa) makes a less "crunchy" sound, and with the short height the sound may be degraded seriously. Therefore, recordings for short alfalfa have a lower signal-to-noise ratio.

Confusion matrices are shown in Table 2 to provide a better idea about which type of events were more frequently confused. From these matrices it may be inferred that the more confusable events in the tests were chew and chew-bites in tall pastures, chew-bites and bites in short alfalfa and chew-bites and chews in short fescue. In general, there were no confusions between chews and bites, denoting that these sounds are quite different.

More information is required for reliable extrapolation of the results of this experiment to cows differing in breed, age and size. Chewing activity of cattle presents pronounced individual differences influenced by physiological state, age and body size [17], but they do not necessarily cause differences in chewing sound.

5 Conclusions

The HMM-based recognition system was capable of automatically segmenting and classifying the masticatory 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. High recognition rates has been obteined with two different pastures and heights. Moreover, the recognizer was able to generalize for all the cases, which denotes robustness in the training algorithms and proposed models.

Automatic acoustic monitoring of ingestive behaviours is also valuable to assess animal welfare in a manner that cannot be achieved with other methods, not even by direct visual observation. Ruminants can chew and bite within a single jaw movement. Thus, mechanical or visual methods cannot fully discriminate these events into simple bites or chew-bites. This study provides a basis for future work on the complete automation of recording, segmentation and classification of masticatory sounds for cattle intake, for behaviour studies and a wide application of the acoustical method.

References

1. Lunguinbuhl, J., Pond, K., Burns, J.: A simple electronic device and computer interface system for monitoring chewing behaviour of stall-fed rumiant animals. Journal Dairy Science **70** (1987) 1307–1312
2. Delagarde, R., Cudal, J., Peyraud, J.: Development of an automatic bitemeter for grazing cattle. Ann Zootech (1999) 329–339
3. Alkon, P., Cohen, Y., Jordan, P.: Towards an acoustic biotelemetry system for animal behaviour studies. Journal Wildl. Manage **53** (1989) 658–662
4. Galli, J., Cangiano, C., Laca, E., Demment, M.W.: The sound of chewing. In et al (eds.), F.P.O., ed.: XX International Grassland Congress: Offered Papers, The Netherlands, Wageningen Academic Publishers (2005) 490
5. Laca, E., Wallis De Vries, M.F.: Acoustic measurement of intake and grazing behaviour of cattle. Grass Forage Sci **55** (2000) 97–104
6. Galli, J., Cangiano, C., Laca, E.: Análisis acústico del consumo de forrajes frescos y secos en novillos. Rev. Arg. Prod. Anim. **23**(1) (2003) 174–175
7. Galli, J., Laforcada, H., Milone, D., Cangiano, C.: Evaluación de un programa de computación para la clasificación automática de sonidos masticatorios. Rev. Arg. Prod. Anim. **25** (2005) 190–191
8. Guerin, F., Milone, D.H., Cangiano, C., Galli, J., Martinez, C.E., Laca, E.: Registro y clasificación de eventos masticatorios de ovinos en pastoreo. In: 7th Argentine Symposium on Computing Technology, 36 JAIIO, Mendoza (2006)
9. Kuc, R.: Introduction to digital signal processing. McGraw-Hill Book Company (1988)
10. Duizer, L.: A review of acoustic research for studying the sensory perception of crisp, crunchy and crackly textures. Trends in Food Science and Thechnology. Elsevier Science Ltd. **12** (2001) 17–24
11. Fukunaga, K.: Introducction to Statistical Pattern Recognition. Academic Press. Werner Rheinboldt, USA (1990)
12. Young, J., Evermann, G., Gales, M., Hain, T., Kershaw, D., Moore, G.: The HTK Book. (2005)
13. Rabiner, L., Juang, B.: Fundamentals of Speech Recognition. Prentice Hall (1993)
14. Huang, X., Acero, A., Hon, H.: Spoken Language Processing: a guide to theory, algorithm and system development. Prentice Hall, USA (2001)
15. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. Proc. IEEE **77**(2) (1989) 257–286
16. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. 2 edn. John Wiley and Sons (1999)

17. 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