Computational Method for Segmentation and Classification of Ingestive Sounds in Sheep

D. H. Milone [a,b,*], H. L. Rufiner [a,b,d], J. R. Galli [c], E. A. Laca [f], C. A. Cangiano [e]

[a] *Laboratorio de Señales e Inteligencia Computacional. Facultad de Ingeniería y Cs. Hídricas, Universidad Nacional del Litoral CC 217, Ciudad Universitaria, Paraje El Pozo, S3000 Santa Fe, Argentina*

[b] *Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina*

[c] *Facultad de Cs. Agrarias, Univ. Nacional de Rosario, Argentina*

[d] *Lab. de Cibernética, Fac. de Ingeniería, Univ. Nac. de Entre Ríos, Argentina*

[e] *EEA Balcarce, Instituto Nacional de Tecnología Agropecuaria, Argentina*

[f] *Department of Plant Science, Univ. of California, Davis*

Preprint submitted to Comp. and Elect. in Agriculture *October 21, 2008*

Abstract

In this work we propose a novel method to analyze and recognize automatically sound signals of chewing and biting. For the automatic segmentation and classification of acoustical ingestive behaviour of sheep the method use an appropriate acoustic representation and statistical modelling based on hidden Markov models. We analyzed 1813 seconds of chewing data from four sheep eating two different forages typically found in grazing production systems, orchardgrass and alfalfa, each at two sward heights. Because identification of species consumed when in mixed swards is a key issue in grazing science, we tested the possibility to discriminate species and sward height by using the proposed approach. Signals were correctly classified by forage and sward height in 67% of the cases, whereas forage was correctly identified 84% of the time. The results showed an overall performance of 82% for the recognition of chewing events.

Key words: Acoustic modeling; Hidden Markov models; Grazing sheep; Ingestive behaviour.

1 Introduction

Accurate measurement of feeding behavior are essential for a reliable management and investigation of grazing ruminants. An indication of animal health and welfare can be obtained by monitoring grazing and rumination activities, because ruminants have a daily chewing requirement to maintain a healthy rumen environment.

* Corresponding author.
Email address: d.milone@ieee.org (D. H. Milone).

Grazing ruminants spend a large part of their lives biting (grasping and severing the herbage in the field) and chewing (grinding of the herbage inside the mouth). They expend quite a lot of energy in eating behavior. When consuming high quality roughage, roughly 10% of the energy content of the feed is consumed in the eating process. On low quality feed, such as wheat straw, that figure jumps to about 25% of energy content (Susenbeth et al., 1998).

Other methods for chewing behavior studies rely on direct observation or on the use of switches and jaw strap adjustment (Stobbs and Cowper, 1972; Penning, 1983; Matsui and Okubo, 1991; Rutter et al., 1997). Direct observation is costly and frequently infeasible. Both direct observation and methods based on jaw movements cannot detect the overlap between chewing and biting. Acoustic biotelemetry has been proposed for animal behavior studies because of the rich information contained in sounds (Alkon et al., 1989) and because sound can be recorded and collected without affecting animal behaviour (Laca et al., 1992; Klein et al., 1994; Nelson et al., 2005).

Acoustic analysis has been proved useful to discriminate a combined chew-bite during a single jaw movement (Laca et al., 1994), to identify some spectral differences of biting and chewing in cattle and it was shown to be a promising method to estimate voluntary intake in different types of feed (Laca and Wallis DeVries, 2000; Galli et al., 2006).

Nevertheless, this method requires further research and development to document the potential of acoustic monitoring of ingestive behavior of animals to yield a consistent automatic decoding of chewing sounds in a variety of conditions. Although discrimination of eating (Nelson et al., 2005; Ungar and Rutter, 2006) and ruminating (Cangiano et al., 2006) sounds appears to be accurate, in the past, sound records have been analyzed manually which is an arduous task. A system for automatic processing and recognition of sound signals is needed to refine and speed up the method.

Automatic speech recognition (ASR) has been an active field of research in the past two decades (Huang et al., 2001). The main blocks of a speech recognizer are: speech signal analysis, acoustic and language modeling. Statistical methods such as hidden Markov models (HMM) have performed well in ASR (Rabiner and Juang, 1986). It is likely that the methods and technologies developed for ASR will be applicable to the analysis of sounds produced by the ingestive behaviour of ruminants. Recently, this type of tools has been used to study and characterize vocal sounds generated by vocalizations of other animals such as red deer (Reby et al., 2006). The objective of our research was to propose a novel method, based on an appropriate signal representation and HMM, to allow the automatic segmentation and classification of bites and chews in sheep grazing a variety of pastures.

The general structure of the proposed system resembles to that of a speech recognition system, where phoneme models are replaced by masticatory sub-events and word models by complete events (such as a chew, bite or combined chew-bite event). As in the speech case, the language model (LM) captures the long-term dependencies and constrains the possible sequences.

This paper was organized as follows. In the next section a brief introduction to hidden Markov models is provided. In Section 3, the sound registration procedure and the data employed for the experiments are presented. Then the statistical model of the acoustic signal is developed and the measures for evaluating its performance were presented. Next, the results of the proposed

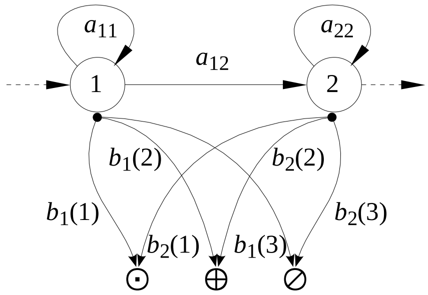

Figure 1. A discrete hidden Markov model with 2 states and 3 output symbols: a_{ij} are the transition probabilities and $b_i(k)$ is the probability of emitting the symbol o_k in state i.

experiments are presented and discussed. Finally the conclusions and suggestions for future work are posed.

2 Hidden Markov models

In this section we will introduce the main concepts behind hidden Markov models, through a simple numerical example. Suppose we want to model sequences of discrete symbols $\mathbf{X} = x_1, x_2, \ldots x_T$, where $x_t \in \mathcal{O}$, the set of possible symbols, and $t \in \mathbb{N}$ stands for the time order in the sequence. For example, with the symbols set $\mathcal{O} = \{\odot, \oplus, \oslash\}$ we can think of the graph of Figure 1 as a generative model for sequences like $\mathbf{X}$. This is an instance of a discrete hidden Markov model (DHMM), with only 2 states ($\mathcal{Q} = \{1, 2\}$), and the 3 symbols as the possible outputs. In a Markov chain, given present state q_t, the next state q_{t+1} is independent of the past states $q_{t-1}, q_{t-2}, \ldots, q_1$. Therefore, the transitions between states can be defined with the matrix of probabilities $A = [a_{ij} = \Pr(q_{t+1} = j | q_t = i)]$, where $i, j \in \mathcal{Q}$. For the generation of sequences, each state i is related to the symbols o_k by an emission distribution $b_i(k) = \Pr(x = o_k | q = i)$. Using all these components, the DHMM can be defined with the structure $\Theta = \{\mathcal{O}, \mathcal{Q}, A, B\}$.

Consider the model Θ with the following probabilities:

$$A = \begin{bmatrix} \frac{9}{10} & \frac{1}{10} \\ 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{1}{10} & \frac{1}{10} & \frac{4}{5} \\ \frac{9}{10} & \frac{1}{20} & \frac{1}{20} \end{bmatrix},$$

where B is the emissions matrix, that specifies the probability of observing each symbol in the actual state. Given the output sequence $\mathbf{X} = \oplus, \oslash, \odot, \odot$, we now ask for the probability of $\mathbf{X}$ given that the model is fully specified. Since we do not know the sequence of states that generated this output sequence, we say that the model is *hidden*, and the states are often referred as latent variables of the model. Thus, to compute the probability of $\mathbf{X}$ given the model Θ, we need to consider all the possible sequences of states and sum up over all the cases (that is, the total probability formula).

The model in Figure 1 is known as a left-to-right HMM, because there are only forward links and self loops (notice that $a_{21} = 0$). In this example, the first state in a sequence will always be the state 1, known as initial state. Similarly, state 2 is the terminal state and all the sequences of states should end with this state. Thus, once the terminal state is reached, the model must observe all the remaining symbols in the same state (that is, $a_{22} = 1$). For a sequence of four symbols, all the possible sequences of states are: $\mathbf{q}^1 = 1 \rightarrow 1 \rightarrow 1 \rightarrow 2$, $\mathbf{q}^2 = 1 \rightarrow 1 \rightarrow 2 \rightarrow 2$, and $\mathbf{q}^3 = 1 \rightarrow 2 \rightarrow 2 \rightarrow 2$. In the first sequence of transitions we have: the emission of symbol $\oplus$ in state $q_1 = 1$, $b_1(2) = \frac{1}{10}$; the transition from the first state to itself at time 2, $a_{11} = \frac{9}{10}$; the emission of symbol $\oslash$, $b_1(3) = \frac{4}{5}$; the second transition $a_{11} = \frac{9}{10}$; the emission of symbol $\odot$, $b_1(1) = \frac{1}{10}$; the third transition, now from state 1 to state 2, $a_{12} = \frac{1}{10}$; and the last emission, of symbol $\odot$ in state $q_4 = 2$, $b_2(1) = \frac{9}{10}$. By using all these probabilities we can obtain $\Pr(\mathbf{X}|\mathbf{q}^1) = \frac{1}{10}\frac{9}{10}\frac{4}{5}\frac{9}{10}\frac{1}{10}\frac{1}{10}\frac{9}{10} \approx 0.0006$. Similarly, we get $\Pr(\mathbf{X}|\mathbf{q}^2) \approx 0.006$ and $\Pr(\mathbf{X}|\mathbf{q}^3) \approx 0.0004$. Then, the probability of the emission sequence given the model is $\Pr(\mathbf{X}) = \Pr(\mathbf{X}|\mathbf{q}^1) + \Pr(\mathbf{X}|\mathbf{q}^2) + \Pr(\mathbf{X}|\mathbf{q}^3) \approx 0.007$.

For the classification tasks we build an HMM for each event to be recognized and then, given an emission sequence of unknown class, we classify it as that corresponding to the most probable model. As it can be seen in the previous example, the probability of the most probable sequence of states is a good approximation to the total probability.

The training or parameter estimation problem remains unaddressed. That is, given a set of sequences of emissions, we are looking for the model probabilities A and B that best fit the data. An intuitive process can be: obtain the best sequences of states for each sequence of emissions and then count the number of transition between states. Thus, probabilities in A can be approximated by the relative frequencies of transitions. In the same way, by counting the times that each state emit a symbol, we can estimate the emission probabilities in B. This algorithm is known as the forced-alignment training and it is based in the fast algorithm proposed by Viterbi (1967). A more complete estimation that uses all the sequences of states (weighted by its probabilities) can be done with the Baum-Welch training method. The forward-backward algorithm provides an efficient way to compute the probabilities for all the sequences and reestimates the parameters in an acceptable processing time for real applications (Huang et al., 1990).

In the case of the acoustic modeling, the sequences of symbols are indeed sequences of spectra. Acoustic models of sounds have evolved from vector quantization with DHMM to more direct models of spectral information with continuous observation density hidden Markov models (CHMM). In the former systems, acoustic features are mapped to a finite set of discrete elements, that is, the outputs of a vector quantizer (Gray, 1984). Thus, it was possible to use DHMM to model sequences of spectral representations. However, in CHMM it is possible to model continuous observation densities in the HMM itself – instead of vector quantization, taking advantage of modeling selected features through Gaussian mixtures (Rabiner and Juang, 1993). To train these models, Liporace (1982) defined the auxiliary function to use in the expectation-maximization (EM) algorithm. He proved that using Gaussian mixtures in the HMM states, this auxiliary function has a unique global maximum as function of the model parameters. The Baum-Welch training uses this EM algorithm and can reach the global maximum given that, as proved in the same work, the sequence of reestimates obtained in produce a monotonic increase in the likelihood of the data given the model.

3 Materials and methods

3.1 Sound registration

Acoustic signals were obtained from a grazing experiment performed at University of California, Davis. The experiment consisted of a factorial of two forages and two plant heights grazed by sheep. The forages were orchardgrass (*Dactylis glomerata*) and alfalfa (*Medicago sativa*) in a vegetative state. Alfalfa and orchardgrass were offered in two plant heights, tall (not defoliated, 29.8 ± 0.79 cm) and short (clipped with scissors to approximately 1/2 the height of the tall treatment, 14.1 $\pm$ 0.79 cm). Pots were firmly attached to a heavy wooden board that held them in place when grazed. We used four tame crossbreed ewes that were 2-4 years old and weighed 85 $\pm$ 6.0 kg.

The order of all combinations of species, plant height and ewe was randomized, and each day (between 12 and 16 h) 8-9 of these combinations (sessions) were observed during six consecutive days. Randomization was restricted such that each of the four combinations of species and plant height and the three ewes were observed at least in one session each day. Animals were fed alfalfa hay and spent the rest of the day in an adjacent yard.

Sounds were recorded using wireless microphones (Nady 155 VR, Nady Systems, Inc., Oakland, California) protected by a rubber foam and placed on the animal's forehead fastened to a halter where the transmitter was attached (see Figure 2). Sound was recorded on the sound track of a digital camcorder.

A watch alarm was set to go off next to the microphone every 10 s as standard sound. Recordings contain various types of natural environmental noises, like bird songs. However, denoising was not applied, signals were fed to the recognizer as recorded in natural environment. Sounds were originally digitized into uncompressed PCM wave format, using a mono channel with a resolution of 16 bits at a sampling frequency of 44100 Hz. Due to the low-frequency nature of signals they were re-sampled to 22050 Hz after processing by an appropriate low-pass filter.

A preliminary comparison of the typical waveforms and spectrograms illustrates the differences between a chew, a bite and a chew-bite of grazing sheep. Figure 3 shows the temporal sound wave and spectrogram of a sequence of chews, bites and chew-bites of grazing sheep. Bites appear as a sequence of

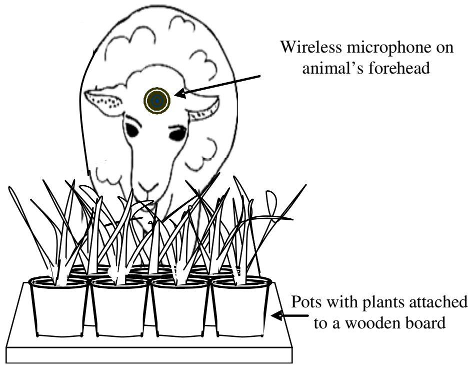

Figure 2. Schematic illustration of the experimental device for sound recording.

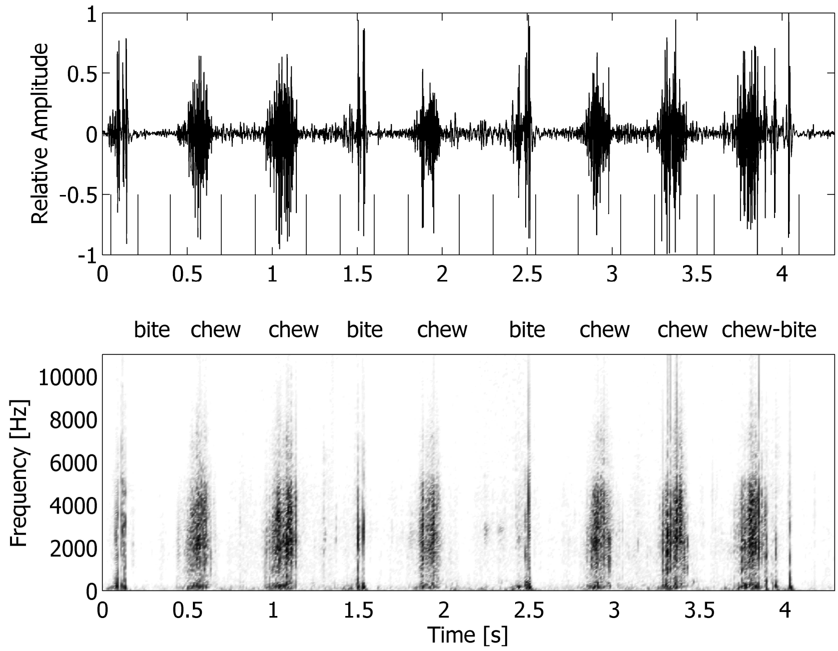

Figure 3. Sound wave (top) and narrow band spectrogram (bottom) of a sequence of chews, bites and chew-bites of grazing sheep.

short bursts of high frequency (Figure 4). Chewing has a relative high energy in the lower half of the spectrum, sustained during a large proportion of its duration (Figure 5). The chew-bite is a composite signal, relatively difficult to be distinguished from the isolated chew signal (Figure 6). Sounds of all events have non-stationary behaviour and their relative frequency contents overlap.

The corpus was formed by the original sound recordings of sheep grazing short alfalfa, tall alfalfa, short orchardgrass and tall orchardgrass. Chews,

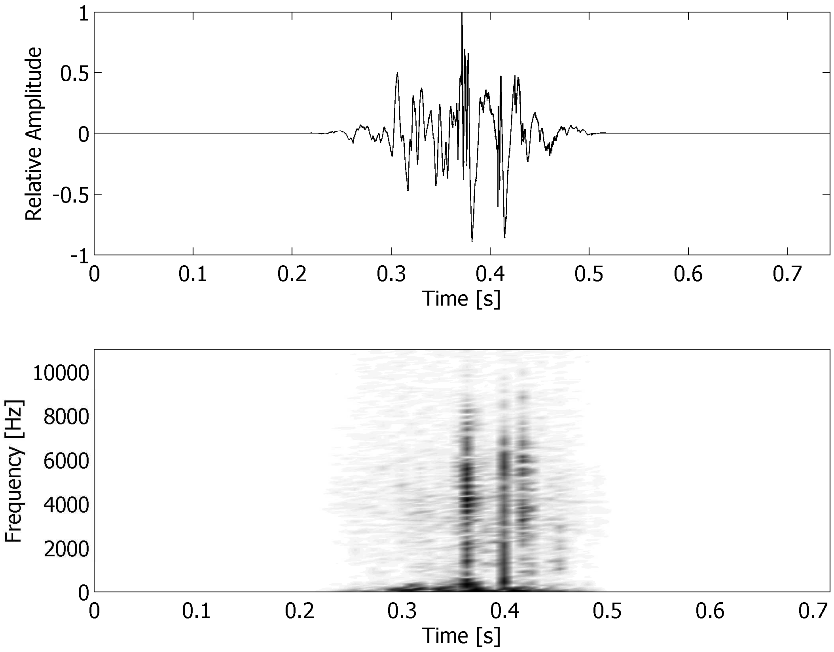

Figure 4. Sound wave (top) and narrow spectrogram (bottom) of an artificially isolated bite of grazing sheep.

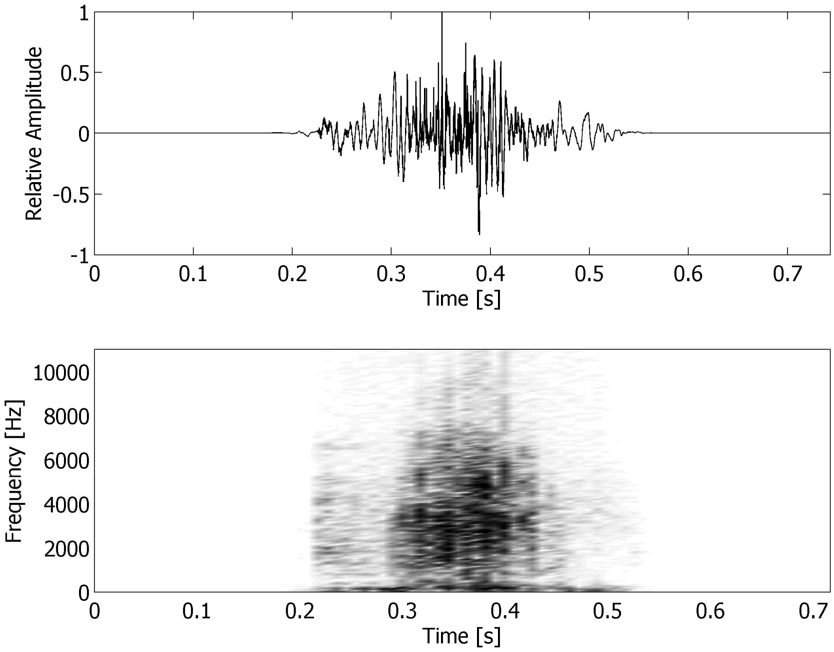

Figure 5. Sound wave (top) and narrow band spectrogram (bottom) of an artificially isolated chew of grazing sheep.

bites and chew-bites were identified and labeled by animal behaviour experts through direct vieweing and listening of the video files. The total lengths of the registered sound data base were 563 and 457 seconds for tall and short

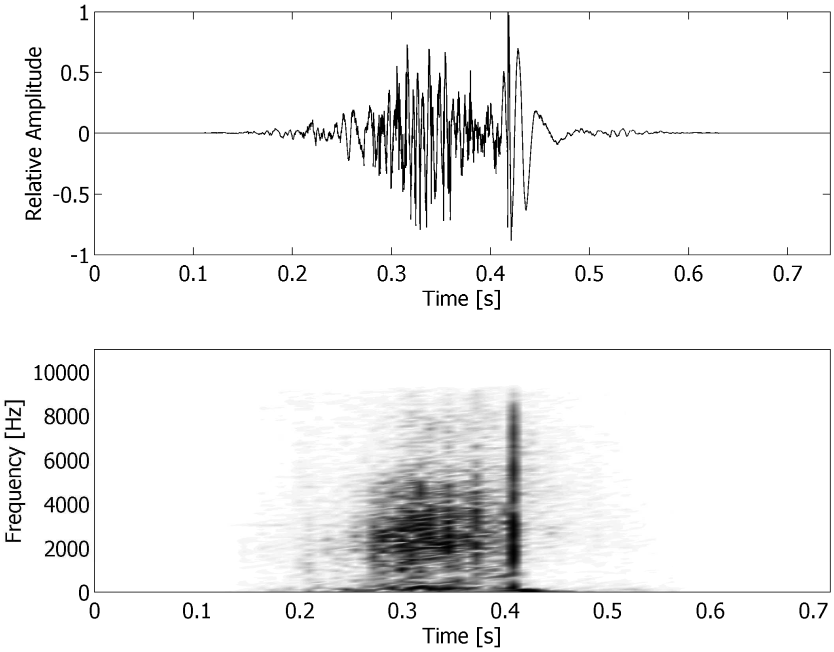

Figure 6. Sound wave (top) and narrow band spectrogram (bottom) of an artificially isolated chew-bite of grazing sheep.

alfalfa and 420 and 373 seconds for tall and short orchardgrass.

3.2 Signal analysis and recognition model

Signal preprocessing consisted of a preemphasis filter and mean subtraction. Because of the non-stationarity of the signal, the next step is to apply short-time analysis techniques. This type of analysis consists in split the whole signal record in short segments called *frames* (Cohen, 1995). These frames are extracted with some periodicity named *step* and have a characteristic time *width*. If the frame width is greater than the step, two consecutive frames will share some samples from the signal record. Each frame was smoothed with a Hamming window to avoid the border effects of the splitting process, and then it was analyzed with several standard spectral estimation techniques, such as: linear prediction coefficients (LPC), linearly spanned filter-bank (FB), log spanned FB, mel FB, cepstrum and mel cepstrum (Oppenheim and Schafer, 1989).

In the recognition model design, each event (*bite, chew* and *chew-bite*) was modeled as a concatenation of sub-events ($bite = [b1 \hookrightarrow b2]$, $chew = [c1 \hookrightarrow c2 \hookrightarrow c3]$ and $chew\text{-}bite = [cb1 \hookrightarrow cb2 \hookrightarrow cb3]$) and, each sub-event as an HMM. Also, one sub-event could have several states and each state could observe one or more frames. The silence was also modeled as a *sil* event [1]. Finally, the individual event models were associated by a language model (LM) into a compound model for every possible sequence of events.

In Fig. 7 we can see the compound model used for the recognition experiments (expanded with more detail for the event *chew*). Each HMM has its states, transition probabilities and probability density functions for the emissions. In our case, gaussian mixtures were used at each state to model the spectral features (i.e., CHMM). We used the above mentioned Baum-Welch algorithm in order to train the HMM transition and observation probabilities.

The LM allows concatenating large sequences of events to model the whole signal record (in a statistical way). In a simple bigram LM, the *a priori* probability that an event occurs given a previous event is used (Jelinek, 1999). For example, a *bite* is frequently followed by a *chew*, thus, the probability for the sequence [*bite chew*] is greater than the probability for the sequence [*bite bite*]. In the same sense the probability for the sequence [*chew chew*] would be greater than the probability for the sequence [*chew bite*]. Figure 8 shows the general bigram LM used in this work.

Once trained, we used the model to recognize unknown sounds by means of standard Viterbi algorithm (Rabiner and Juang, 1993). All the experiments

[1] In some experiments the *sil* event was modeled as a sub-event of the others events.

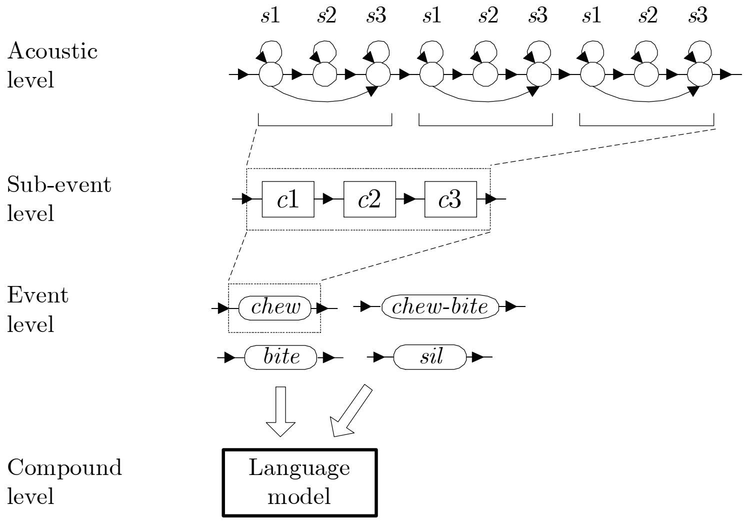

Figure 7. Schematic diagram of the compound model used for the recognition experiments. The different states represent the events at the acoustic level, while individual HMMs have been used to represent the sub-event level. Each sub-event model was merge into an event model by a dictionary and all this events are associated in a compound model by a language model.

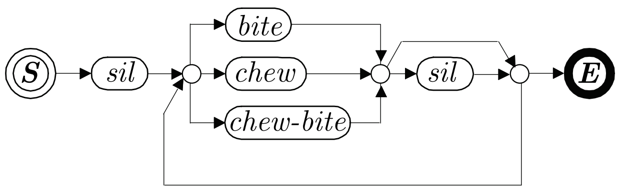

Figure 8. General language model used for the experiments. S means "start" and E means "end".

were implemented using the HTK toolkit [2]. The configuration details will be given in the next section.

The before mentioned compound model was used for the experiments with only one species and height of pasture. In this case the model had three possible output classes of events for the sequence (*bite, chew* and *chew-bite*). It was a *pasture dependent* model (for one species and one height). Four different pasture dependent models were trained: tall alfalfa (M_{TA}), short alfalfa (M_{SA}), tall orchardgras (M_{TO}) and short orchardgrass (M_{SO}).

In order to construct the *pasture independent* model (M_{ind}) we followed a

[2] `http://htk.eng.cam.ac.uk`

similar way but using twelve events classes (and silence): one *bite, chew* and *chew-bite* different model for every one of each four pasture conditions. For example for tall alfalfa we had $bite_{TA}$, $chew_{TA}$ and $chew\text{-}bite_{TA}$ events, for short alfalfa we had $bite_{SA}$, $chew_{SA}$ events and $chew\text{-}bite_{SA}$, and so on. All these events were associated in an overall model, extending the LM to include 12 events.

3.3 Recognition performance measures

In order to achieve statistically representative results, all the tests were conducted according to the averaged leave-k-out method (Michie et al., 1994). The complete data set was randomly split 10 times in train/test sets, using the 80% of total signals for training and the remaining 20% for recognition.

For each test partition, the performance was evaluated by using the recognition rate and the recognition accuracy, respectively as follows:

$$C_j = \frac{\sum_{i=1}^{N_j} T_{ji} - D_{ji} - S_{ji}}{\sum_{i=1}^{N_j} T_{ji}} \qquad A_j = \frac{\sum_{i=1}^{N_j} T_{ji} - D_{ji} - S_{ji} - I_{ji}}{\sum_{i=1}^{N_j} T_{ji}}$$

where:

N_j : total number of sequences in test partition j

T_{ji} : total number of events in the sequence i of the test partition j

D_{ji} : deleted events in the recognized sequences i of the test partition j

S_{ji} : substituted events in the recognized sequences i of the test partition j

I_{ji} : inserted events in the recognized sequences i of the test partition j

In all the counts, silence events are ignored and the final results are computed as the average among all test partitions. To obtain the D_{ji}, S_{ji} and I_{ji} counts, a dynamic programming alignment between the hand labeled reference sequence (Ref_{seq}) and the recognized sequence (Rec_{seq}) was performed (Young et al., 2000). An example of the recognizer output is reproduced here with the corresponding hand-labeled reference:

Ref_{seq}:	[*bite*	*chew*	*chew*		*chew*	*chew*	*bite*	*chew*	*chew*	]
Rec_{seq}:	[*bite*	*chew*	*chew*	$\underline{bite}$	*chew*	*chew*	$\underline{chew}$	*chew*		]

The second bite from the recognized sequence is an insertion and the fifth chew is a substitution. The last event from the reference sequence has not been recognized so it is a deletion. From this example it can be seen that A_j is a much more exigent measure since it considers insertions while C_j does not. A_j is very useful to make decisions about the overall system performance, because if I_{ji} is greater than T_{ji}, C_j can be high (near 100%) while A_j will be negative. However, once a system has reached a reasonable A_j (say, 80%) then C_j can be used to select the best system configuration.

We also used the *confusion matrix* as another method to evaluate classification performance. In confusion matrices, each column represents a predicted event class, while each row represents the actual class. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes.

4 Experiments and Results

For the feature extraction the best recognition rates among parametric techniques were for LPC. For non-parametric techniques best results were for linearly spanned FB. We employed 10 linearly spanned bands between 0 and 2000 Hz for the FB and 20 coefficients for LPC. In addition, an estimation of the derivatives and total energy in the spectral estimation was used. Frame width ranged from 10 ms to 100 ms in steps of 10 ms and overlapping ranged from 0% to 75% in steps of 25%. Best results were obtained with a frame width of 20 ms and an overlapping of 25%.

In the tuning of the HMM architecture, we first made tests using one model per event, where the number of emitting states ranged from 2 to 8. These results were useful as a baseline for the following stage. Based on visual inspection of the sound signal, chews and chew-bites were modeled by three sub-events, bites with two and silence with only one. This division in sub-events is justified by the different spectral evolution between ingestive events but not for the silence that should remain almost stationary.

Using the Gaussian means of the trained models, one can build the estimated spectra that HMM holds in each state. This could provide some insights about the internal model representation of spectral signal dynamics, remarking the important characteristics that it takes into account. As an example we used the M_{TA} trained. The spectral dynamic for bite is modeled by two sub-events ($b1$ and $b2$) each with three states (Figure 9). Chew is modeled by three sub-events ($c1$, $c2$ and $c3$) each with three states (Figure 10). Chew-bite is also modeled by three sub-events ($cb1$, $cb2$ and $cb3$) each with three states (Figure 11). Bite spectra for the first state of $b1$ has an important peak around 1 kHz, alternating between narrow and broad-band spectra for the rest of the states. It also finishes the last state of $b2$ with a peaked spectrum again. If we carefully inspect Figure 4, this is compatible with the sequence of short duration bursts (naturally broad-band) and low energy regions (which presents

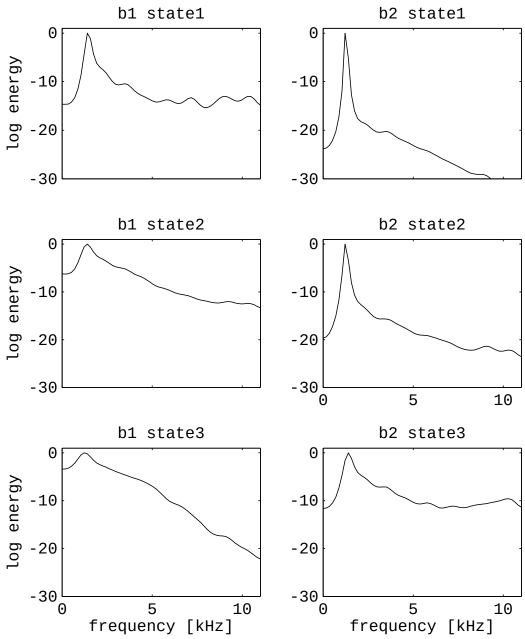

Figure 9. Spectra of bite estimated by the tall alfalfa model after the training process, computed from the Gaussian means of LPC parameters of each model state. Columns: sub-events; Rows: states.

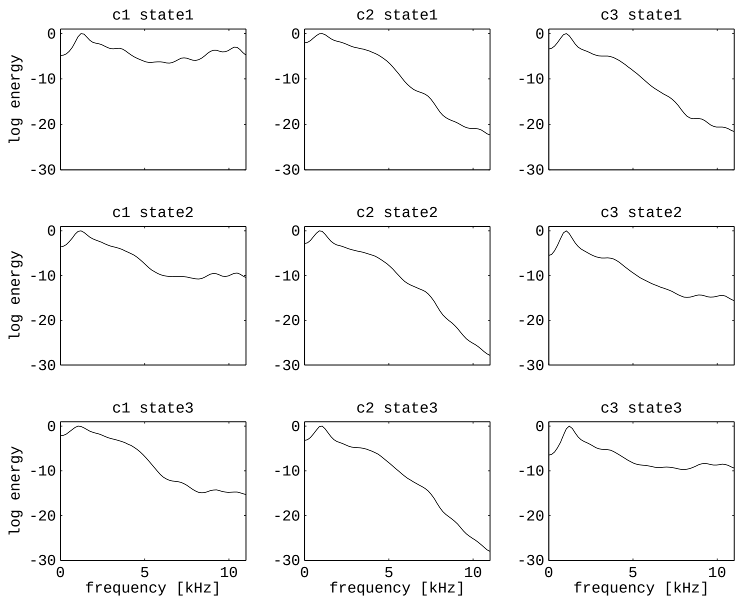

Figure 10. Spectra of chew estimated by the tall alfalfa model after the training process, computed from the Gaussian means of LPC parameters of each model state. Columns: sub-events; Rows: states.

relative concentration of energy around 1 kHz) displayed in it. Chew spectra varies from one with a relative flatness for $c1$ to one more concentrated around the bottom half in $c2$, finishing with spectra in $c3$ similar to $c1$ (but in reverse

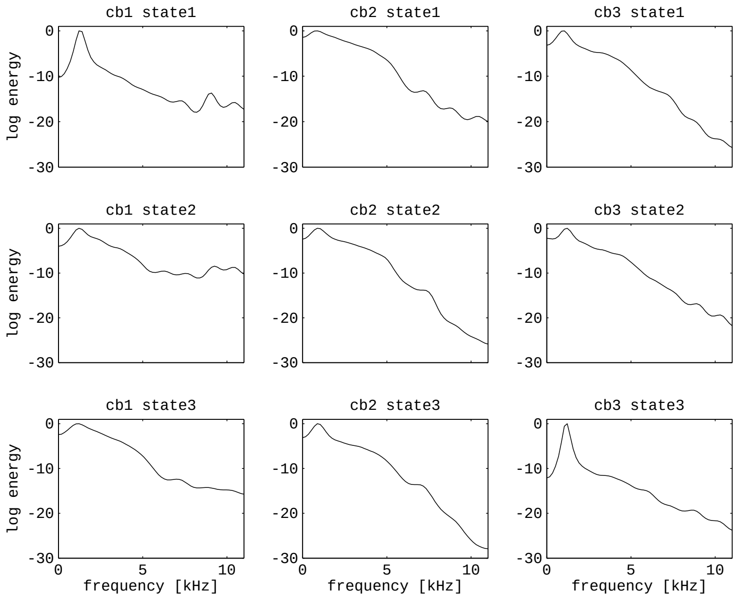

Figure 11. Spectra of chew-bite estimated by the tall alfalfa model after the training process, computed from the Gaussian means of LPC parameters of each model state. Columns: sub-events; Rows: states.

order). This is clearly related with the example of chew presented in Figure 5. As expected, spectrum of the first sub-events of chew-bite were similar to those of chew while the last (*cb*3 state 3, Figure 11) was similar to those of bite. This is associated with the composite signal nature observed in Figure 6.

Our preliminary results (not shown here for brevity) revealed that, for all the cases, the models that used LPC as parameterization performed better than those that used FB. Also, the use of more than one model per event obtained higher recognition rates, compared to the baseline of one model per event.

After extensive trials, the best recognition results were obtained with a window of 20 ms and an overlapping 15 ms, using 3 emitting states per HMM. In this case, the system yielded an average A_j of 81.60% and an average C_j of 89.47%. This analysis was carried out using the tall alfalfa files and M_{TA} model, and it set the basic system configuration that was employed afterwards.

Table 1
Recognition results classifying pasture species and height, by relabeling M_{ind} model (see details in text).

	Average A$_j$ (%)	Average C$_j$(%)
$M_{ind}^{dir-sil}$ species and height	51.59	58.11
$M_{ind}^{sub-sil}$ species and height	61.63	66.53
$M_{ind}^{sub-sil}$ species only	80.54	83.78

We will first present the detailed results of the pasture independent model M_{ind}. In this case we employed the previous described configuration and the extended LM for 12 events. We used two different models for these experiments: modelling the sil event directly ($M_{ind}^{dir-sil}$) or as a sub-event of the others events ($M_{ind}^{sub-sil}$): $chew = [c1 + c2 + c3 + sil]$ and $bite = [b1 + b2 + sil]$. As explained before, the model was trained with the different pastures conditions.

The grazing behaviour and diet selection by sheep are influenced by the sward height and the relative species content of pastures (Illius et al., 1992). Consequently it would be useful for reseachers to obtain an accurate classification of ingestive sounds for different species and different heights within species. In order to obtain different recognition results once trained we re-labelled the final classes or events in the model. The first case was for the recognition of species and height of pasture, that is a four classes problem. For example for tall alfalfa class we replace $bite_{TA} \rightarrow TA$, $chew_{TA} \rightarrow TA$ and $chew\text{-}bite_{TA} \rightarrow TA$. The second required result was the recognition of species only, that is a two classes problem. For example for alfalfa class we replace $bite_{TA} \rightarrow A$, $chew_{TA} \rightarrow A$, $chew\text{-}bite_{TA} \rightarrow A$, $bite_{SA} \rightarrow A$, $chew_{SA} \rightarrow A$ and $chew\text{-}bite_{SA} \rightarrow A$. The results for this problems are shown in Table 1. When dealing with mixtures of pastures, the system turned on recognizing the species and height of pasture with a performance of 66.53%, while on the recognition of only the species the

Table 2
Confusion matrix for classifying bites, chews and chew-bites, by relabeling M_{ind} model (values in parenthesis are %).

	bite's	*chew*'s	*chew-bite*'s
bite's	438 (58)	160 (21)	154 (21)
chew's	182 (4)	4658 (89)	380 (7)
chew-bite's	90 (13)	222 (31)	402 (56)

result grew up to 83.78%.

Another case is to use $M_{ind}^{sub-sil}$ to recognize the event without taking into account the pasture conditions (i.e. for bite class we replace $bite_{TA} \rightarrow bite$, $bite_{SA} \rightarrow bite$, $bite_{TO} \rightarrow bite$ and $bite_{SO} \rightarrow bite$). The results are displayed in Table 2. Overall classification performance of events was 82%, where we obtained 58% for bites, 89% for chews and 56% for chew-bites (see confusion matrix diagonal, Table 2). Chews were classified as chew-bites in 7% of the cases, chew-bites were partially misclassified as bites in 13% and chews in 31%.

In the pasture dependent models, where only one species and height were considered, recognition values were generally higher than the previous ones. The results for tall alfalfa $M_{TA}^{sub-sil}$ were 74, 96 and 61% of bites, chews and chew-bites, respectively. Similar results were obtained for short alfalfa $M_{SA}^{sub-sil}$ (68, 94 and 49 %). Tall orchardgrass $M_{TO}^{sub-sil}$ resulted in 66, 90 and 39% and short orchardgrass $M_{SO}^{sub-sil}$ performed slighty worse with 18, 77 and 74%, probably due to a relative lower signal to noise ratio for this type of sounds.

5 Discussion

To our knowledge this is the first time an automatic recognition of ingestive chewing of ruminants is done and it extends the use of HMM beyond vocalizations studies in wild animals (Clemins et al., 2005; Reby et al., 2006).

The systems was effective and robust for the automatic segmentation and classification of chewing behaviour in sheep. It speed-up the processing of data and leaving the real time factor from hours to minutes, including the analysis of other useful variables such as energy of chews (Galli et al., 2006) while (Laca and Wallis DeVries, 2000) took nearly of 120 hours to manage nearly 1000 bites and chews of steers.

The individual recognition models for each species have better performance than using a unique overall model. These last ones had a high error rate mainly because of event substitution. Nevertheless, they could be useful for preliminary segmentation of signals with unknown species and height of pasture.

The values obtained for window width and overlapping are indicators of the signal stationarity, in the sense that its spectral characteristics (LPC or the spectrum itself) do not show significant variation in the interval, so one frame can be distinguished from another using a spectral distance measure. The 3-state models have demonstrated to be sufficient to model human speech phonemes, so we consider reasonable that best results have been obtained with models of few states. Carefully inspection of Figures 9, 10 and 11 seems to indicate that the proposed number of states and sub-events allows the model to correct follow the spectral dynamic of the events. A priori, chewing sounds do not exhibit the complexity of speech, at least when considering the sound generation mechanisms and the information content of both signals.

The surrounding noise such as animal vocalization and scratching against the ground are difficult to filter out automatically. The overall system performance is being altered because this sounds are generally classified as masticatory events (they are not strictly silence) and this yields an error by insertion. However, these sounds can be discarded by a previous stage because in general are of short duration and low energy compared to those of the masticatory events. Studies to determine if these sounds are very frequent must be made, and a model for these events could be added to the system in order to improve the general performance. This fact does not imply a major deviation in the overall signal energy thus does not represent a problem when using the chew energy for dry matter intake prediction (Laca and Wallis DeVries, 2000; Galli et al., 2006). However, they represent an error when analyzing grazing behaviour because they introduce a bias in the chews per bite ratio. Several denoising techniques are also available, which could be used to improve signal quality previously the starting of the recognition process.

The compound jaw movements, namely chew-bites and already detected in cattle by Laca and Wallis DeVries (2000) were acoustically confirmed in sheep and their spectra automatically recognized by trained HMM.

Since much of the acoustic signal generated by mechanical interaction of teeth and food during occlusion is transmitted by bone conduction, the direct attachment of the microphone to the forehead of bovine (Laca et al., 1994) or head of mule deer (Nelson et al., 2005) picks up a wider range of sound than a free-standing (collar–mounted) microphone, and can more easily pick up the vibrations associated with mastication. The location is unobtrusive and proven acceptable in controlled applications but for grazing extensive conditions could be exposed to serious damage. The ear canal is one another possible place to locate the microphone as a more secure and insulated position as already proved in human (Amft et al., 2005). Furthermore, our system can be used for real-time monitoring, at short distances, with wireless microphones. For an implementation in a wide area, we are planning to incorporate a transmission system over the cellular network or to use a recording system for several days and then to process the signals in an standard personal computer. Alternatively, a digital signal processor may be integrated with the microphone in the forehead of each animal, but this may be a quite expensive solution.

6 Conclusions and future work

We conclude that the automatic segmentation and classification of masticatory sounds by sheep is possible by modelling the acoustical signals with continuous hidden Markov models. Models were tuned for optimal performance using a compound model with different levels of analysis, from the acoustic of sub-events to the long-term dependence given by the intake language model. This study provides a basis for future work on the complete automation of recording, segmentation and classification of masticatory sounds for intake and grazing animal behaviour studies and for a wide application of the acoustical method.

The ultimate goal of a system that precisely and reliably determines the type and amount of food that the animal consumed is far. However, considering the rate at which new commercial technologies become available, and the development and analysis of larger sound data-bases, we can visualize an automated acoustic monitoring system for ingestive behaviour in ruminants.

The novel approach we used has potential to be used no only as a technique to automatically record and classify ingestive sounds, but also as a new way to describe ingestive behaviour and to relate it to animal and forage characteristics. We hypothesize that the parameters for the language models are synthetic descriptors of ingestive behavior that have the ability to integrate the characteristics of feeding bouts into a few numbers. These numbers, such as transition probabilities between behaviors and components of events could be used to gain insight in the ingestion process. Automatic acoustig 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. Chewing is a fundamental behavior for the maintenance of rumen function and animal well-being because it supplies the rumen with saliva, enzymes and buffering compounds. Acoustic monitoring provides the most accurate quantification of chewing, and could be developed into a routine method to monitor animals such as dairy cows that are subject to the stresses of extremely high productivity.

7 Acknowledgements

This work was carried out with the financial support of ANPCyT and UNL, under Projects PICTO-UNL No 36214 and CAI+D No 012-72, and of INTA under Project PNLEC No 1101. The authors would like to thank Fernando Guerin for his collaboration in the visual inspection and labeling of sound signals and for models scripts debugging.

References

Alkon, P., Cohen, Y., Jordan, P., 1989. Towards an acoustic biotelemetry system for animal behaviour studies. J. Wildlife Manage 53, 658–662.

Amft, O., Stäger, M., Lukowicz, P., Tröster, G., 2005. Analysis of chewing sounds for dietary monitoring. In: Proceedings of the 7th International Conference on Ubiquitous Computing. Tokyo, Japan, pp. 56–72.

Cangiano, C., Galli, J., Laca, E., 2006. El uso del sonido en el análisis de la rumia. Visión Rural 62, 23–24.

Clemins, P., Johnson, M., Leon, K., Savage, A., 2005. Automatic classification and speaker identification of african elephant (loxodonta africana) vocalizations. Journal of the Acoustical Society of America, 956–963.

Cohen, L., 1995. Time Frequency Analysis: Theory and Applications. Prentice-Hall.

Galli, J., Cangiano, C., Demment, M., Laca, E., 2006. Acoustic monitoring of chewing and intake of fresh and dry forages in steers. Animal Feed Science and Technology 128, 14–30.

Gray, R., 1984. Vector quantization. IEEE Acoustics Speech and Signal Processing Magazine 4, 4–29.

Huang, X., Acero, A., Hon, H.-W., May 2001. Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Prentice Hall PTR.

Huang, X. D., Ariki, Y., Jack, M. A., 1990. Hidden Markov Models for Speech Recognition. Edinburgh University Press.

Illius, A. W., Clark, D. A., Hodgson, J., 1992. Discrimination and patch choice by sheep grazing grass-clover swards. Journal of Animal Ecology 61, 183–194.

Jelinek, F., 1999. Statistical Methods for Speech Recognition. MIT Press, Cambrige, Masachussets.

Klein, L., Baker, D., Purser, D., Zacknich, A., Bray, A., 1994. Telemetry to monitor sounds of chews during eating and rumination by grazing sheep. Proceeding of the Australian Society of Animal Production 20, 423.

Laca, E., Ungar, E., Seligman, N., Ramey, M., Demment, M., 1992. An integrated methodology for studying short-term grazing behaviour of cattle. Grass and Forage Science 47, 81–90.

Laca, E. A., Ungar, E. D., Demment, M. W., 1994. Mechanisms of handling time and intake rate of a large mammalian grazer. Applied Animal Behaviour Science 39, 3–19.

Laca, E. A., Wallis DeVries, M. F., 2000. Acoustic measurement of intake and grazing behaviour of cattle. Grass and Forage Science 55, 97–104.

Liporace, L. A., 1982. Maximum likelihood estimation for multivariate stochastic observations of Markov chains. IEEE Trans. Information Theory 28 (5).

Matsui, K., Okubo, T., 1991. A method for quantification of jaw movements suitable for use on free-ranging cattle. Applied Animal Behaviour Science 32, 107–116.

Michie, D., Spiegelhalter, D., Taylor, C., 1994. Machine Learning, Neural and Statistical Classification. Ellis Horwood, University College, London.

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

Oppenheim, A. V., Schafer, R. W., 1989. Discrete-Time Signal Processing. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Penning, P. D., 1983. A technique to record automatically some aspects of grazing and ruminating behaviour in sheep. Grass and Forage Science 38, 89–96.

Rabiner, L. R., Juang, B. H., 1986. An introduction to hidden Markov models. IEEE Acoustics Speech and Signal Processing Magazine 3 (1), 4–16.

Rabiner, L. R., Juang, B. H., 1993. Fundamentals of Speech Recognition. Prentice-Hall.

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

Rutter, S., Champion, R., Penning, P., 1997. An automatic system to record foraging behaviour in freeranging ruminants. Applied Animal Behaviour Science 54, 185–195.

Stobbs, T. H., Cowper, L. J., 1972. Automatic measurement of the jaw movements of dairy cows during grazing and rumination. Tropical Grasslands 6, 107–111.

Susenbeth, A., Mayer, R., Koehler, B., Neumann, O., 1998. Energy requirement for eating in cattle. Journal of Animal Science 76, 2701.

Ungar, E., Rutter, S., 2006. Classifying cattle jaw movements: Comparing iger behaviour recorder and acoustic techniques. Applied Animal Behaviour Science 98, 11–27.

Viterbi, A. J., 1967. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. IEEE Transactions on Information Theory 13, 260–269.

Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V., Woodland, P., 2000. HMM Toolkit. Cambridge University, http://htk.eng.cam.ac.uk.