Automatic detection of cow/calf vocalizations in free-stall barn

Stavros Ntalampiras*, Andrea Pezzuolo§, Silvana Mattiello‡, Monica Battini‡ and Marta Brščić†

*Department of Computer Science, University of Milan, Milan, Italy

§Department of Land, Environment, Agriculture and Forestry, University of Padova, Legnaro, Italy

‡Department of Agricultural and Environmental Sciences, University of Milan, Milan, Italy

†Department of Animal Medicine, Production and Health, University of Padova, Legnaro, Italy

Email: stavros.ntalampiras@unimi.it

Abstract—Precision livestock farming dictates the use of advanced technologies to understand, analyze, assess and finally optimize a farm's production collectively as well as the contribution of each single animal. This work is part of a research project wishing to steer the dairy farms' producers to more ethical rearing systems. To study cow's welfare, we focus on reciprocal vocalizations including mother-offspring contact calls. We show the set-up of a suitable audio capturing system composed of automated recording units and propose an algorithm to automatically detect cow vocalizations in an indoor farm setting. More specifically, the algorithm has a two-level structure: a) first, the Hilbert follower is applied to segment the raw audio signals, and b) second the detected blocks of acoustic activity are refined via a classification scheme based on hidden Markov models. After thorough evaluation, we demonstrate excellent detection results in terms of false positives, false negatives and confusion matrix.

Keywords—Precision livestock farming; cow/calf vocalization; acoustic signal processing, vocalization detection; audio pattern recognition.

I. Introduction

The actions of modern dairy farmers are closely monitored for various reasons ranging from growing consumers' awareness and production costs to issues like antibiotic resistance and environmental impact [1], [2]. At the same time, globally, there is an increase of animal protein demand and a higher consciousness with consumers seeking for sustainable ethically-produced food [3]. As a response, there is increasing interest in a) keeping dairy cows with their calves during the milk feeding period to reduce distress associated with separation (animals emit characteristic vocalizations), and b) alternative dairy production systems that allow continued cow-calf contact and promote natural behaviors with positive effects on animal welfare.

Overall, the project aims at investigating the acceptability and feasibility of potential alternative systems with a prolonged cow-calf contact and to assess humans' ability to recognize the emotional state of cattle from vocalizations emitted in different contexts with special attention on early separation.

Increased availability of Automated Recording Units (ARUs) enables automatic monitoring of dairy farms which could be a useful tool in all above-mentioned directions. Such monitoring is typically carried out via acoustic and/or visual modalities [4], [5]. Since animal vocalizations convey emotional, physiological and individual information [6], this paper focuses on the acoustic modality. Conveniently, acoustic emissions are practically indifferent to both occlusions and light conditions while they can be captured with affordable sensors.

The analysis of animal vocalizations for precision livestock farming purposes comprises an active research field as it may benefit the current production systems in terms of resource efficiency, environmental impact, consumer's trust, transparency, and security [7] beyond improving animal welfare assessment [8], [9]. The work described by Bishop et al. in [10] includes automated sound detection system for sheep, cattle, and dogs. Several authors have focused on vocalizations produced during oestrus: Röttgen et al. [11] describe such a system recording in group-housed dairy cattle, while an exploration of the optimal formants for acoustic detection of cow's oestrus is presented by Lee et al. [12] and Cung et al. [13]. Moving on, Green et al. [14] studied cattle mother-offspring contact calls and more specifically how these encode individual-identity information. Furthermore, they studied the associations existing between cattle vocalizations with positive (oestrus and anticipation of feed) and negative (denied feed access and upon both physical and physical and visual isolation from conspecifics) welfare contexts. Interestingly, the authors of [15] examined natural calls captured by crossbred beef cows and their calves under undisturbed conditions. It was evident that low frequency calls were emitted when the cows were in close proximity to their calves during the initial 3-4 weeks postpartum with mouth closed or only partially open. On the contrary, high frequency calls were observed in case the cows were not allowed visual contact with their calves. From another point of view, Psĕnka et al. [16] highlight the fact that cattle are exposed to various unnatural noises in existing farm environments which are mainly the result of mechanical equipment facilitating farm operations (e.g. fans, water circulation, etc.). They studied the noise level exposure during their day routine and associated it
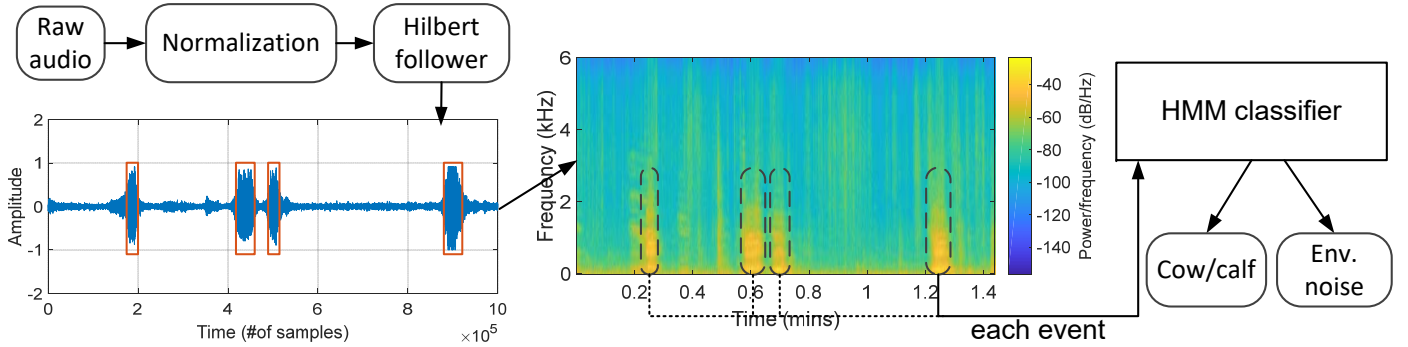
Fig. 1. The proposed method for detection of cow-calf vocalizations. After raw signal normalization, the Hilbert follower is employed and finally the detections are refined by an HMM-based classification scheme.

with their hearing sensitivity at selected frequencies.

This work is concentrated on the entire range of cattle mother-offspring contact as well as their individual calls. To analyze the acoustic characteristics related to animals' emotional state, the primary step is capturing the associated cow/calf vocalizations as they occur in a dairy farm. This paper outlines an algorithm enabling automatic detection of cow/calf vocalization in continuous recordings. A hierarchical structure is designed where: a) the first layer detects segments of potential acoustic activity based on the Hilbert follower, and b) the second layer includes a *detection-by-classification* scheme based on Hidden Markov Model (HMMs) refining the previously detected chunks. Finally, we present the recording protocol and report excellent results in terms of false positive and false negative rates on data recorded over 12h. The proposed framework is able to successfully detect cow/cattle vocalizations and discard data associated with background noise reasonably fast, while processing raw recordings made in a real dairy farm environment. At the same time, the hierarchical detection framework is flexible and can be adapted to detect other species as long as the respective data becomes available.

II. HIERARCHICAL ACOUSTIC ACTIVITY DETECTION

This section describes all modules of the proposed algorithm, i.e. raw audio pre-processing, Hilbert follower and HMM-based acoustic activity detection.

A. Audio signal pre-processing and Hilbert follower

As the frequency range in which the vocalizations of cows/cattle might appear is not assumed to be known, we considered the entire spectrum without filtering. The signal's mean value is subtracted and its gain normalized so that possible calibration problems of the microphone and/or DC-offsets were eliminated.

Acoustic activity detection is carried out using the Hilbert follower which is responsible for segmenting the captured audio signal following the distinctive shape of the animal vocalization's envelopes. The isolated segments are employed

during the learning stage as shown in Fig. 1. The Hilbert follower is a particularly important part of the detection algorithm as it was proven able to discard a massive quantity of data not including cow/calf vocalizations.

We briefly describe its operation: let us denote the acquired discrete-time audio signal $y(n)$, where n is the time index, and the Hilbert follower as $\mathcal{H}$. When $\mathcal{H}$ is applied on $y(n)$, i.e.

$$y^h(n) = \mathcal{H}(y(n)) \tag{1}$$

we obtain a series of complex numbers which comprises the *analytic* signal of $y(n)$ [17], [18]. The analytic signal is comprised by a real part $y(n)$ containing the initial datastream, and an imaginary part, $y_i(n)$ including the Hilbert transform of $y(n)$. Subsequently, the envelope $e(n)$ of the time-domain signal is computed as follow:

$$e(n) = \sqrt{y^h(n) \otimes \widehat{y^h(n)}} \tag{2}$$

where $\widehat{y^h(n)}$ comprises the conjugate of $y^h(n)$ and $\otimes$ the component-wise multiplication operator.

During acoustic activity detection, the envelope computed as in Eq. 2 is checked against a threshold θ. Whenever a given sample of $y(n) > \theta$, it is considered as *acoustic activity*. On the contrary, i.e. when $y(n) \leq \theta$ sample n is regarded as *non-activity*. Interestingly, since the thresholding process is directly associated with the time-domain signal, we can naturally extract time instances and segments of acoustic activity as shown in Fig. 1. Importantly, such time-stamped information can be very useful when initializing the HMMs, which is the next stage of the proposed methodology, since it removes the need for manual annotation.

B. HMM-based activity detection refinement

The segments identified as acoustic activity by the Hilbert follower, are further refined for potential errors using a more computationally intensive algorithm, i.e. a classifier using two HMMs. Chunks of acoustic activity might include environmental noise events (e.g. metallic object impact); towards discarding them, we designed a binary classifier.
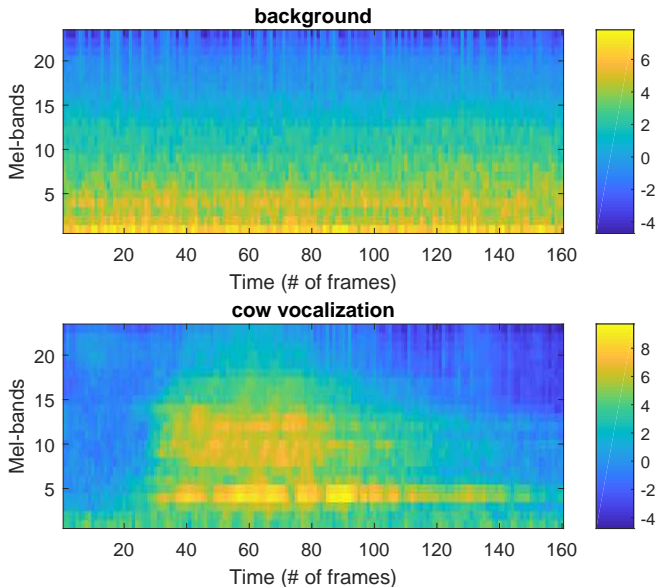
Fig. 2. Mel-spectrograms of background noise (top) and a cow vocalization (bottom).

Fig. 3. The placement of a Song Meter SM4 Acoustic Recorder.

In brief, we extracted 13 Mel-frequency cepstral coefficients out of the detected segments. To this end, each segment is windowized into stationary frames of 30ms overlapped by 20ms which is a reasonable choice for generalized sound recognition tasks [19], [20]. For each window, the short-time Fourier transform is computed and then passed via a triangular Mel-scale filterbank. Afterwards, the logarithm of powers at each of the Mel frequencies is calculated and finally, the discrete cosine transform is applied to decorrelate the data. The first 13 coefficients are retained including the $0-th$ one encompassing information directly associated with the energy of the signal. Mel-spectrograms of background noise and cow vocalization are illustrated in Fig. 2.

The HMM-based classification scheme follows the class-specific logic as described in [21], [22]. Accordingly, two HMMs are constructed to represent each sound class, i.e. *cow/calf vocalization* and *environmental noise*, denoted $\mathcal{H}^c$ and $\mathcal{H}^n$ respectively. A predefined set of possible numbers of states was identified based on the problem specifics while the distribution of each state is approximated by a Gaussian mixture model of diagonal covariance suitably elaborating on the extracted MFCCs. During operation, the features extracted out of the novel signal are fed to both HMMs and each one outputs a log-likelihood representing the probability that the given sequence of features was emitted by them. Finally, classification is carried out based on the *maximum log-likelihood* criterion.

III. VOCALIZATION DETECTION EXPERIMENTS

This section describes the a) recording protocol that was followed, b) the figures of merit, c) the parameterization of the proposed method, as well as d) the experimental results.

A. Indoor farm space and recording protocol

The present research was carried out in a private dairy farm located near Treviso in Northeast Italy. The farm was characterized by a free-stall system and housing 40 lactating cows. Continuous 24/7 acoustic farm monitoring was carried out via three Song Meter SM4 Acoustic Recorders[1]. Placement of one recording device is shown in Fig. 3. Two ARUs were placed in the primary area and one in a protected space dedicated to deliveries. Care was taken during microphone calibration to choose a suitable gain given the ARU position with respect to its distance from the animals. The sampling rate was set to 24kHz which well covers the target animal vocalizations. We used 16-bit coding and all files are stored in standard WAV format.

The recording period started on 19/9/2019 at approximately 12:00am and spans four weeks in order to capture a) various birth events and the respective mother-offspring vocalizations which took place in a separate space close to the primary one and b) the entire gamut of the possible interferences.

Naturally, all activities taking place in the farm environment were recorded. Hence, captured audio is influenced by factors that could not be controlled such as rain, snow, wind, noise, interferences from background activities, etc. All such events were kept in the dataset as our target is to process the raw recordings without any significant efforts on manual annotation [23]. Indeed, data labeling was carried out only for the first 6 hours of recordings (environmental noise, and animal vocalizations) based on the outcome of Hilbert follower as described in section II-A. This part of the dataset was used for constructing the HMMs.

Last but not least, the research supported by the present project does not include information allowing person identification in any aspect (physical, mental, cultural, societal, etc.), since human speech is considered an interference and, hence,

---

[1]https://www.wildlifeacoustics.com/products/song-meter-sm4

TABLE I. THE CONFUSION MATRIX (IN %) OBTAINED BY THE APPLICATION OF HMM-BASED CLASSIFICATION. THE ACHIEVED AVERAGE RECOGNITION RATE IS 70.1%.

Method \ Figure of merit	FP (%)	FN (%)
Hilbert follower	1.3	9.3
Hilbert follower+HMM	-	**4.3**

TABLE II. THE CONFUSION MATRIX (IN %) OBTAINED BY THE APPLICATION OF HMM-BASED CLASSIFICATION. THE ACHIEVED AVERAGE RECOGNITION RATE IS 97.9%.

Presented \ Responded	*Background noise*	*Cow/cattle*
Background noise	**95.7**	4.3
Cow/cattle	-	**100**

discarded. Finally, representative Mel-spectrograms are shown in 2.

B. Figures of merit and parameterization

In order to assess the detection capabilities of the proposed algorithm in both levels, i.e. Hilbert follower with and without the HMM-based classifier, we used the following suitably-defined figures of merit:

- *False positive* (FP): it comprises the number of times a cow vocalization is detected without being presents (percentage).
- *False negative* (FN): it comprises the number of times an existing cow vocalization is not detected (percentage).

Moreover, a confusion matrix was used to evaluate the classification ability of the proposed HMM scheme.

After early experimentations, threshold θ was set equal to the mean signal energy augmented by the double of its standard deviation. As regards to the HMM classifiers, the final number of states and Gaussian functions was selected by exhaustively searching the sets $N_{st} = \{3, 4, 5\}$ and $N_G = \{2, 4, 8, 16, 32\}$. The number of iterations of k-means cluster initialization and Baum-Welch algorithms are bounded by 50 and 25 respectively. Finally, the probability threshold between subsequent iterations was 0.001 [24]. It should be mentioned that HMMs were trained and validated using events in the first 6h of the recordings and tested on the following 6h.

C. Results

The application of Hilbert follower discards approximately 94% of the recorded data which, after a macroscopic view of the datastreams, manages to retain all animal activity. Table I includes the FP and FN rates respectively. As we can see, they are kept within very low levels ($< 10\%$) while the addition of the HMM scheme leads to 4.3% of FNs without any FP, i.e. a significant performance boost. The respective confusion matrix is shown in Table II.

Given that the events of interest are very sparse in time, automated detection of cow/calf vocalizations becomes highly relevant as it reduces the need for manual labor as well as data storage. At the same time, such accurate event detection can be used to timestamp the incoming data so that the corresponding visual data can be easily retrieved and analyzed in order to investigate the cause/etiology behind the detected vocalizations.

IV. CONCLUSIONS

This article analyzed an algorithm enabling automated detection of animal vocalizations in an indoor farm setting to be used for PLF purposes. Interestingly, the proposed solution is able to reach excellent results even though it operates under conditions influenced by highly non-stationary environmental noise events.

In the future, timestamped cow/calf vocalizations including mother-offspring contact calls will be combined with the synchronized video recordings in order to contextualize each vocalization into a specific *positive* or *negative* context and associate given behavioural state(s). Overall, we hope to sensitize producers towards more ethical rearing systems and to stimulate controversial discussions of the future sustainability of the dairy sector and related dairy-type meat productions.

REFERENCES

[1] A. Brombin, A. Pezzuolo, and M. Brščić, "Are we ready for the big change in the dairy production system?" *Research in Veterinary Science*, vol. 126, pp. 17–19, Oct. 2019. [Online]. Available: https://doi.org/10.1016/j.rvsc.2019.08.006

[2] R. Geers and F. Madec, Eds., *Livestock production and society*. Wageningen Academic Publishers, Mar. 2006. [Online]. Available: https://doi.org/10.3920/978-90-8686-567-3

[3] R. Mandel, H. Whay, E. Klement, and C. Nicol, "Invited review: Environmental enrichment of dairy cows and calves in indoor housing," *Journal of Dairy Science*, vol. 99, no. 3, pp. 1695–1715, Mar. 2016. [Online]. Available: https://doi.org/10.3168/jds.2015-9875

[4] C.-S. Chen and W.-C. Chen, "Research and development of automatic monitoring system for livestock farms," *Applied Sciences*, vol. 9, no. 6, p. 1132, Mar. 2019. [Online]. Available: https://doi.org/10.3390/app9061132

[5] K. Wurtz, I. Camerlink, R. B. D'Eath, A. P. Fernández, T. Norton, J. Steibel, and J. Siegford, "Recording behaviour of indoor-housed farm animals automatically using machine vision technology: A systematic review," *PLOS ONE*, vol. 14, no. 12, p. e0226669, Dec. 2019. [Online]. Available: https://doi.org/10.1371/journal.pone.0226669

[6] M. P. Mcloughlin, R. Stewart, and A. G. McElligott, "Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring," *Journal of The Royal Society Interface*, vol. 16, no. 155, p. 20190225, Jun. 2019. [Online]. Available: https://doi.org/10.1098/rsif.2019.0225

[7] R. S. Alonso, I. Sittón-Candanedo, Ó. García, J. Prieto, and S. Rodríguez-González, "An intelligent edge-IoT platform for monitoring livestock and crops in a dairy farming scenario," *Ad Hoc Networks*, vol. 98, p. 102047, Mar. 2020. [Online]. Available: https://doi.org/10.1016/j.adhoc.2019.102047

[8] J. M. Watts and J. M. Stookey, "Vocal behaviour in cattle: the animal's commentary on its biological processes and welfare," *Applied Animal Behaviour Science*, vol. 67, no. 1-2, pp. 15–33, Mar. 2000. [Online]. Available: https://doi.org/10.1016/s0168-1591(99)00108-2

[9] S. Ferrari, R. Piccinini, M. Silva, V. Exadaktylos, D. Berckmans, and M. Guarino, "Cough sound description in relation to respiratory diseases in dairy calves," *Preventive Veterinary Medicine*, vol. 96, no. 3-4, pp. 276–280, Sep. 2010. [Online]. Available: https://doi.org/10.1016/j.prevetmed.2010.06.013

[10] J. C. Bishop, G. Falzon, M. Trotter, P. Kwan, and P. D. Meek, "Livestock vocalisation classification in farm soundscapes," *Computers and Electronics in Agriculture*, vol. 162, pp. 531–542, Jul. 2019. [Online]. Available: https://doi.org/10.1016/j.compag.2019.04.020

[11] V. Röttgen, P. C. Schön, F. Becker, A. Tuchscherer, C. Wrenzycki, S. Düpjan, and B. Puppe, "Automatic recording of individual oestrus vocalisation in group-housed dairy cattle: development of a cattle call monitor," *animal*, vol. 14, no. 1, pp. 198–205, Aug. 2019. [Online]. Available: https://doi.org/10.1017/s1751731119001733

[12] J. Lee, S. Zuo, Y. Chung, D. Park, H. Chang, and S. Kim, "Formant-based acoustic features for cow's estrus detection in audio surveillance system," in *2014 11th IEEE AVSS*, Aug 2014, pp. 236–240.

[13] Y. Chung, J. Lee, S. Oh, D. Park, H. H. Chang, and S. Kim, "Automatic detection of cow's oestrus in audio surveillance system," *Asian-Austr. Journ. of Animal Sciences*, vol. 26, no. 7, pp. 1030–1037, Jul. 2013. [Online]. Available: https://doi.org/10.5713/ajas.2012.12628

[14] A. Green, C. Clark, L. Favaro, S. Lomax, and D. Reby, "Vocal individuality of holstein-friesian cattle is maintained across putatively positive and negative farming contexts," *Scientific Reports*, vol. 9, no. 1, Dec. 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-54968-4

[15] M. P. de la Torre, E. F. Briefer, T. Reader, and A. G. McElligott, "Acoustic analysis of cattle (bos taurus) mother–offspring contact calls from a source–filter theory perspective," *Applied Animal Behaviour Science*, vol. 163, pp. 58–68, Feb. 2015. [Online]. Available: https://doi.org/10.1016/j.applanim.2014.11.017

[16] M. Pšenka, M. Šístková, Š. Mihina, and R. Gálik, "Frequency analysis of noise exposure of dairy cows in the process of milking," *Research in Agricultural Engineering*, vol. 62, no. No. 4, pp. 185–189, Nov.

2016. [Online]. Available: https://doi.org/10.17221/4/2015-rae

[17] V. Bargmann, "Irreducible unitary representations of the lorentz group," *The Annals of Mathematics*, vol. 48, no. 3, p. 568, Jul. 1947. [Online]. Available: https://doi.org/10.2307/1969129

[18] I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Applied Acoustics*, vol. 80, pp. 1–9, Jun. 2014. [Online]. Available: https://doi.org/10.1016/j.apacoust.2014.01.001

[19] S. Ntalampiras, "A classification scheme based on directed acyclic graphs for acoustic farm monitoring," in *2018 23rd IEEE FRUCT*, Nov 2018, pp. 276–282.

[20] ——, "Automatic analysis of audiostreams in the concept drift environment," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, Sep. 2016, pp. 1–6.

[21] S. Ntalampiras and I. Potamitis, "A statistical inference framework for understanding music-related brain activity," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 275–284, May 2019.

[22] S. Ntalampiras, "A novel holistic modeling approach for generalized sound recognition," *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 185–188, Feb 2013.

[23] S. Ntalampiras, D. Arsić, M. Hofmann, M. Andersson, and T. Ganchev, "PROMETHEUS: heterogeneous sensor database in support of research on human behavioral patterns in unrestricted environments," *Signal, Image and Video Processing*, vol. 8, no. 7, pp. 1211–1231, Jun. 2012. [Online]. Available: https://doi.org/10.1007/s11760-012-0346-9

[24] S. Ntalampiras, "Generalized sound recognition in reverberant environments," *Journal of the Audio Engineering Society*, vol. 67, no. 10, pp. 772–781, Oct. 2019. [Online]. Available: https://doi.org/10.17743/jaes.2019.0030