KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

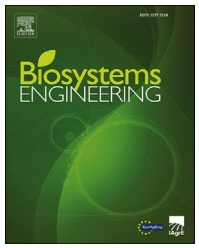
Citation/Reference	Lenn Carpentier, Daniel Berckmans, Ali Youssef, Dries Berckmans, Toon van Waterschoot, Dayle Johnston, Natasha Ferguson, Bernadette Earley, Ilaria Fontana, Emanuela Tullo, Marcella Guarino, Erik Vranken, and Tomas Norton **Automatic cough detection for bovine respiratory disease in a calf house** *Biosystems Engineering*, vol. 173, pp. 45-56, Sep. 2018 (Special Issue on Engineering Advances in Precision Livestock Farming).
Archived version	Author manuscript: the content is identical to the content of the submitted paper, but without the final typesetting by the publisher
Published version	https://doi.org/10.1016/j.biosystemseng.2018.06.018
Journal homepage	https://www.sciencedirect.com/science/journal/15375110
Author contact	toon.vanwaterschoot@esat.kuleuven.be *+ 32 (0)16 321788*
IR	ftp://ftp.esat.kuleuven.be/pub/SISTA/vanwaterschoot/reports/18-149.pdf

(article begins on next page)

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110



Special Issue: Engineering Advances in Precision Livestock Farming

Research Paper

Automatic cough detection for bovine respiratory disease in a calf house



Lenn Carpentier [a], *Daniel Berckmans* [a], *Ali Youssef* [a], *Dries Berckmans* [b], *Toon van Waterschoot* [c], *Dayle Johnston* [d], *Natasha Ferguson* [d], *Bernadette Earley* [d], *Ilaria Fontana* [e], *Emanuela Tullo* [e], *Marcella Guarino* [e], *Erik Vranken* [a,f], *Tomas Norton* [a,*]

[a] *Division Animal and Human Health Engineering, Department of Biosystems, KU Leuven, Kasteelpark Arenberg 30, 3001 Heverlee, Belgium*

[b] *SoundTalks NV, Kapeldreef 60, 3001 Heverlee, Belgium*

[c] *Department of Electrical Engineering (ESAT-STADIUS/ETC), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium*

[d] *Animal and Bioscience Research Department, AGRIC, Teagasc, Grange, Dunsany, Co., Meath, Ireland*

[e] *Department of Health, Animal Science and Food Safety, Università Degli Studi di Milano, Via Celoria 10, 20133, Milan, Italy*

[f] *Fancom Research, Fancom BV, Wilhelminastraat 17, 5981 XW Panningen, The Netherlands*

ARTICLE INFO

Article history:
Published online 6 July 2018

Keywords:
Precision Livestock Farming
Bioacoustics
Sound analysis
Cough
Calves

In calf rearing, bovine respiratory disease (BRD) is a major animal health challenge. Farmers incur severe economic losses due to BRD. Additional to economic costs, outbreaks of BRD impair the welfare of the animal and extra expertise and labour are needed to treat and care for the infected animals. Coughing is recognised as a clinical manifestation of BRD. Therefore, the monitoring of coughing in a calf house has the potential to detect cases of respiratory infection before they become too severe, and thus to limit the impact of BRD on both the farmer and the animal. The objective of this study was to develop an algorithm for detection of coughing sounds in a calf house. Sounds were recorded in four adjacent compartments of one calf house over two time periods (82 and 96 days). There were approximately 21 and 14 calves in each compartment over the two time-periods, respectively. The algorithm was developed using 445 min of sound data. These data contained 664 different cough references, which were labelled by a human expert. It was found that, during the first time period in all 3 of the compartments and during the second period in 2 out of 4 compartments, the algorithm worked very well (precision higher than 80%), while in the 2 other cases the algorithm worked well but the precision was less (66.6% and 53.8%). A relation between the number of calves diagnosed with BRD and the detected coughs is shown.

© 2018 IAgrE. Published by Elsevier Ltd. All rights reserved.

* *Corresponding author.*
E-mail address: tomas.norton@kuleuven.be (T. Norton).

https://doi.org/10.1016/j.biosystemseng.2018.06.018
1537-5110/© 2018 IAgrE. Published by Elsevier Ltd. All rights reserved.

Nomenclature

P	Audio power
x	Amplitudes of the audio samples
K	Number of audio samples
M	Median frequency (Hz) or time (s)
P1	10th percentile frequency (Hz) or time (s)
P2	90th percentile frequency (Hz) or time (s)
IPR	Interpercentile range frequency (Hz) or time (s)
C	Spectral centroid
X	Energies of the frequency bins
Wf_L	Number of frequency bins
S	Spectral spread
n	Normalised spectral energy
H	Spectral entropy
Fl	Spectral flux

1. Introduction

Bovine respiratory disease (BRD) is a multifactorial disease that is driven by complex interactions of factors associated with the environment, the pathogen, the animal, and management practices (Edwards, 2010; Potter & Aldridge, 2010a). In the dairy and beef sectors, BRD is considered an economically important disease internationally (Johnston et al., 2017; Tennant, Ives, Harper, Renter, & Lawrence, 2014). Once an outbreak occurs, it often leads to morbidity, and mortality in feedlot and veal calves, and in dairy calf to beef/dairy replacement rearing systems (Edwards, 2010; Healy et al., 1993; Pardon et al., 2013; Snowder et al., 2007).

Detecting the disease at an early stage, in order to provide early treatment, requires swift observation of the presence of clinical signs associated with the disease. However, this depends on the farmer's judgement and experience. Given that typical cattle farm sizes have increased significantly over the years, the level of quality attention livestock are receiving is diminishing (Berckmans, 2004). Another problem is that, in practice, farmers and veterinarians tend to underestimate signs of BRD morbidity during their clinical examinations (Potter & Aldridge, 2010b; White & Renter, 2009).

Clinical signs associated with BRD are cough, nasal discharge and tachypnoea (Ozkanlar et al., 2012; Potter & Aldridge, 2010b). Cough sounds can thus be used as a biomarker for BRD in calves. For example, for infection with the parainfluenza-3 virus, coughing often tends to be the first visible clinical sign (Allan, Pirie, Selman, & Snodgrass, 1978) making it a very suitable biomarker for swift detection of outbreaks. An advantage of monitoring bioacoustics, such as cough sounds, using microphones is that it is performed non-invasively, and hence does not influence the animals' normal behaviour. Another advantage is that one microphone can monitor many individuals, making such systems affordable to farmers. However, a single microphone is unable to identify the origin of the emitted sound, hence is unable to identify the individual that coughed.

Van Hirtum, Aerts, Berckmans, Moreaux, and Gustin (1999) first published a study on cough detection in pigs, which was followed by additional studies refining the pig cough detection algorithms (Chedad et al., 2001; Chung et al., 2013; Exadaktylos, Silva, Aerts, Taylor, & Berckmans, 2008). However, these studies are limited to laboratory conditions where the individual sound events were manually extracted. The ratio of cough to non-cough events is therefore more balanced, which does not correspond with real-life situations where there are much more non-cough events. More recently, a commercial tool showed it was possible to detect infected pigs 2–12 days before a farmer or veterinarian (Berckmans, Hemeryck, Berckmans, Vranken, & van Waterschoot, 2015).

Recently, Ferrari et al. (2010) and Vandermeulen et al. (2016) studied the application of sound monitoring techniques for calf cough detection. Ferrari et al. (2010) characterised cough and metallic sounds, which are frequently encountered due to the use of metal fences, gates, and racks in calf housing facilities, and investigated acoustic differences between them. The authors showed that there was a significant difference between the duration, fundamental frequency, and amplitude of both sounds and hence they concluded it was possible to discriminate cough sounds from metal rack sounds using these features.

During the development of an algorithm to automatically detect cough sounds, a reference dataset of cough sounds is needed. This dataset is obtained by manually observing and annotating sound recordings (Aerts, Jans, Halloy, Gustin, & Berckmans, 2005; Tullo, Fontana, & Guarino, 2013). Studies under lab conditions result in high quality sounds as the recording device is placed close to the animal (Chung et al., 2013; Guarino, Jans, Costa, Aerts, & Berckmans, 2008) and sounds were recorded of each individual animal (Chung et al., 2013; Exadaktylos et al., 2008). However, in a commercial environment the animals will be monitored in a group and the recording device will be further away. This will result in less clear sound events and overlapping sounds.

Vandermeulen et al. (2016) developed an algorithm for the automatic detection of coughing sounds in calves with a focus on more real-life environments. Similar to the algorithm of Van Hirtum et al. (1999) and Exadaktylos et al. (2008), the Euclidian distance between the power spectral density of a specific sound and a reference dataset of coughs was used for classification. However, to cope with changing environments, the reference dataset needed to be calibrated in order for the algorithm to perform well when going to another compartment. The main problem is that this calibration step required coughs to be manually labelled when enrolling the algorithm in another environment, which is time consuming and difficult, if not impossible, to implement in real-farm situations (Kashiha et al., 2013; Oczak et al., 2013). Techniques that remove the need for this step need to be developed in order to achieve a technology that can work in realistic environments.

Instead the objective is to use general features which describe the characteristics of the cough event independent of the environment it is recorded in. Therefore, features which are less susceptible to noise are explored (Cortopassi, 2006).

Given the previous rationale, the aim of the present study is to develop an algorithm that distinguishes cough sounds from other sounds found in commercial calf rearing facilities. A key innovation of this study is a new approach that obviates the need for calibrated reference labels during monitoring.

Additionally, this study makes use of new labelling approach for the development of the algorithm which takes the quality of the reference sound into account.

2. Material and methods

2.1. Animals and housing

The study was conducted at the Teagasc Beef Research Centre, Grange, Co. Meath in Ireland, coordinates 53° 30′ N, 6° 40′ W. Calves were housed in four adjacent compartments which measured 6.77 m by 4.56 m (a schematic overview can be seen in Fig. 1). The compartments were all open on the front side with three solid walls on the other sides and a galvanised mono-pitch roof and sawdust bedded floors. Sound was recorded using a sound acquisition system comprising a microphone and sound card (described in detail in Section 2.2). The microphone was mounted above the centre of the floor space of the compartment (Fig. 1) at a height of about 2.75 m above the floor.

Sound data were recorded continuously for two different batches. The first batch consisted of 84 bull calves of which 46 were Holstein-Friesian (H-F) and 38 were Jersey (J) calves. The mean age and mass (±the standard deviation) of the calves on arrival for batch 1 were 19 (±8) days and 41.4 (±8.4) kg, respectively. The first batch was recorded from March 21st 2013 until June 11th 2013. For the second batch, 55 predominantly H-F bull calves were enrolled, with mean age and mass on arrival of 26.5 (±8.6) days and 40.9 (±8.3) kg, respectively. The second batch was recorded from March 11th 2015 until June 15th 2015.

The 139 animals used in the study were clinically healthy at the start of the study. The calves were first acclimatised to the electronic feeding system (Foster-Tecknik SA 2000, Engen, Germany) and afterwards housed in mixed groups. Table 1 shows the details about the calves in the different groups at the start of the study for batch 1 and batch 2.

The experimental licence (B100/2869) for the experiments performed during this study was obtained from the Irish Department of Health and Children. Experimental protocols met all criteria imposed by the Cruelty to Animals Act (Ireland 1876, as amended by European Communities regulations 2002 and 2005) and the European Community Directive 86/609/EC.

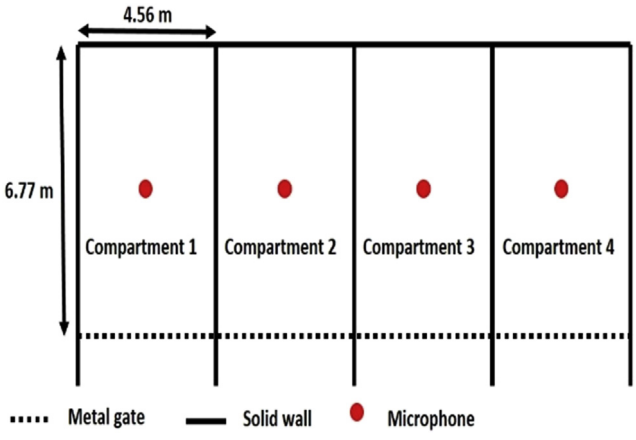


Fig. 1 – Schematic overview of the calf house.

2.2. Sound acquisition

Each calf compartment was equipped with a sound acquisition system (SoundTalks NV). The system contained a sound card (ESI MAYA 44), which was mounted on a fanless embedded board, and a condenser microphone (Behringer C4). Microphones were phantom powered and were connected to the sound card using balanced audio, so the system itself could just be plugged into the power grid. Monaural sounds were recorded with a sampling frequency of 22.05 kHz and a 16-bit resolution. This gave frequency information up to 11.025 kHz (=Nyquist frequency) which was considered enough to detect the sound events of interest. To protect the equipment from the harsh environment, the embedded board was installed in a sealed box and the microphone was protected by a flexible plastic cover.

Sounds were continuously monitored over the time the calves were in the house. During this study 17,280 h of data were recorded. The sound recordings were split into chunks of 5 min and stored on an external hard drive in Waveform Audio File Format (WAV) files, which are uncompressed audio files, so no information was lost due to compression. In total more than 200,000 recordings were saved on the hard drives, which resulted in 2.6 TB of data stored. As the compartments are located next to each other, very loud sounds from neighbouring compartments were picked up by the microphone.

2.3. Labelling

In order to develop an algorithm to detect cough events, a reference for cough events was needed. This reference was given by audio-visual inspection of the sound recordings. Human observers, who each inspected different files, annotated the beginning and end of cough events in a label track, and the observation was performed using an audio-visual representation of the sound recordings (Aerts et al., 2005; Tullo et al., 2013, 2017). Software was used that could replay and visualise sounds using spectrograms for the detection of coughing events. Depending on the observer, either Adobe® Audition™ CS6 software or Audacity® recording and editing software version 2.1.1 were used to label the data. In addition to the detection of coughing sounds, labelling was used to score the quality of measured data. This novelty was introduced because some coughs overlapped with other events or were unclear. The label was a number between 1 and 5, where 1 indicates a very unclear cough and 5 a very clear cough. Coughs with a label greater than or equal to three were considered as coughs which the algorithm should be able to detect.

2.4. Blood analysis as a gold standard

Blood samples were taken from the calves around the weaning period, more specifically at 14, 6, and 3 days before weaning and 0, 1, 3, 8, and 14 days after weaning. Blood was collected using jugular venepuncture and stored in 6 ml

Table 1 – Details about the calves in the different compartments at the start of the study.

		Compartment 1	Compartment 2	Compartment 3	Compartment 4
First batch (March 21 to June 11, 2013)	Number of calves and breed	13 J and 8 H-F	10 J and 11 H-F	8 J and 13 H-F	7 J and 14 H-F
	Mean age, days (±SD)	31.4 (10.0)	28.8 (6.7)	23.7 (9.5)	22.1 (8.0)
	Mean mass, kg (±SD)	39.3 (8.5)	39.4 (7.4)	43.9 (8.7)	40.9 (8.7)
Second batch (March 11 to June 15, 2015)	Number of calves and breed	14 H-F	14 H-F	14 H-F	13 H-F
	Mean age, days (±SD)	15 (2.5)	14 (4.9)	14 (3.1)	13 (2.6)
	Mean mass, kg (±SD)	51.3 (4.9)	42.9 (4.9)	51.6 (4.6)	43.8 (5.1)

K_3Ethylenediaminetetraacetic acid tubes (Vacuette, Cruinn Diagnostics, Ireland). Haematological analysis of neutrophil profiles was performed using an ADVIA 2120 analyser (ADVIA 2120, Bayer Healthcare, Siemens, UK). A calf was diagnosed with BRD when the blood neutrophils exceeded the normal reference value of 4000 μl^{-1} (Jones & Allison, 2007).

2.5. *Algorithm for cough detection*

The algorithm developed for cough detection in calves can be split into four parts: pre-processing, event selection, feature calculation, and classification into cough and non-cough sounds. In the pre-processing step, the raw sound data were cleaned by removing background noise. Pre-processing in this study was based on spectral subtraction similar to the power spectral subtraction filter used in the work of Amrulloh, Abeyratne, Swarnkar, Triasih, and Setyati (2015), which was based on minimum statistics (Martin, 2001). The next step, the event selection, focusses on selecting the beginning and end of possible cough events. In these intervals, different feature variables are calculated. Based on these feature variables the sound events are then classified as either cough or non-cough. The algorithm was developed using a commercial software package (MATLAB 2014b, The MathWorks Inc., Natick, MA, 2000).

2.5.1. *Event selection*

After pre-processing the sounds, the event selection algorithm selects events in the sound data which are possible cough events. This step annotates the beginning and end of the events, converting the sound recordings to a set of individual sound events. Further processing is then performed on this set of selected sound events. Therefore, a good selection of the beginning and end will facilitate the classification algorithm. The objective in this step is to select as many real cough events as possible with a good annotation of beginning and end.

Coughing events contain more energy than the background noise, hence detection based on energy envelope can be performed. Examples of event selection methods used in other bioacoustics studies are based on sudden changes in the standard deviation of the recorded sound (Barry, Dane, Morice, & Walmsley, 2006; Vandermeulen et al., 2016), and the envelope of the energy (Aydin, Bahr, & Berckmans, 2015; Aydin et al., 2014; Fagerlund, 2007).

The algorithm to detect the sound events in this paper is based on the energy content. The energy envelope is obtained by summating the filtered spectrogram over the frequency range. An adaptive baseline is obtained by calculating the 10th percentile - this value was determined *ad hoc* by maximising the overlap between the algorithm output and the labelled sound. This overlap was calculated as the length of the intersection of both labels divided by length of the union of both labels. A sound event was defined as a time period that exceeds the baseline. The beginning was defined as the point where the energy exceeded the adaptive threshold and the end was defined as the first point after a beginning point where the energy dropped below the adaptive threshold (Fig. 2). Subsequently, the beginning was corrected based on the spectral flux of the selected interval. This new beginning, together with the end, define the event's start and end points. The beginning and end point are only accepted if the duration between them is larger than 0.1 s. This restriction was added so that selected features could be calculated for all events.

2.5.2. *Feature calculation*

In this step, feature values were calculated for each sound event defined by the event selection algorithm. Ferrari et al. (2010) described that a cough can be distinguished from a metal sound based on the amplitude, duration, and fundamental frequency of the sound. However, in real-life conditions, those features will be affected by noise, distance from the calf to the microphone and directionality of the microphone. Because of this, more robust features were taken which are less susceptible to noise and varying conditions. These features are derived from the waveform or the spectrogram of the sound signal (Fig. 3).

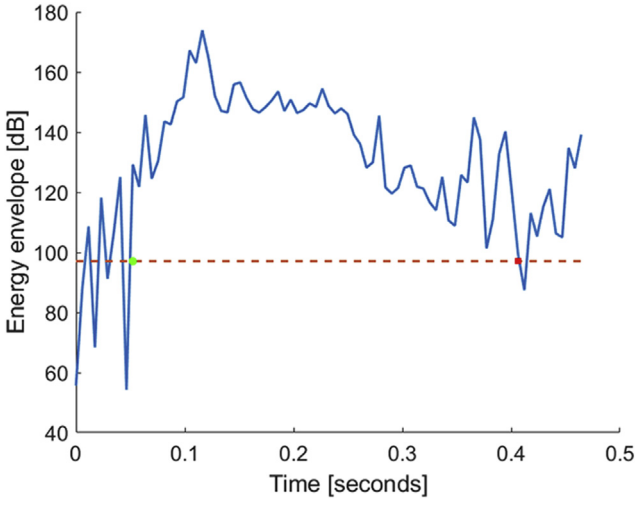


Fig. 2 – Example of event selection: the round point represents the beginning of an event, when the energy exceeds the baseline. The square point represents the end of the event, when the energy drops below the baseline.

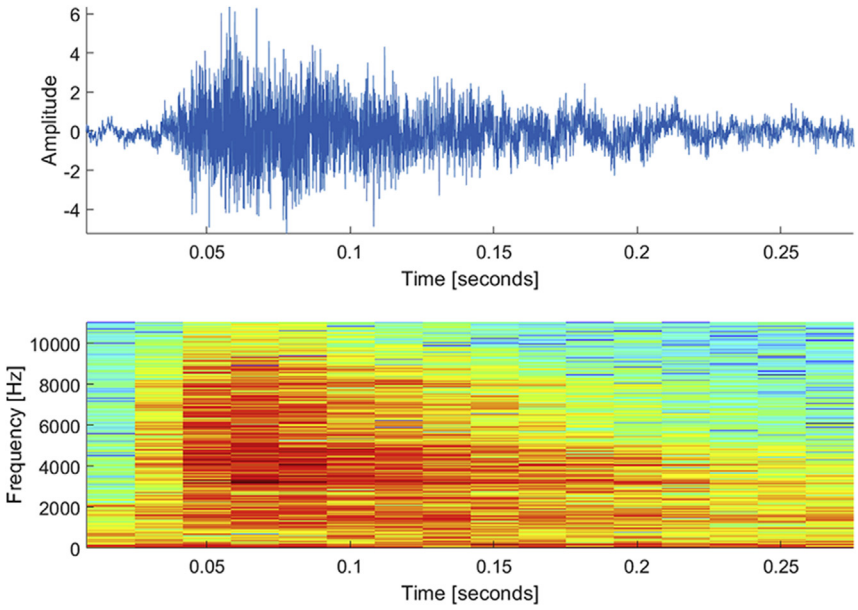


Fig. 3 – Waveform (top figure) and spectrogram (bottom figure) of a calf cough.

Similar to the studies of Ferrari et al. (2010) and Vandermeulen et al. (2016), the duration of an event is explored as possible feature. The duration is calculated as the end time minus the beginning time of the selected event and is expressed in seconds. Additionally to the duration of the total event, the time it took for a cough to get to maximum energy (rising time) and the time from the maximum to the end of the event (decaying time) were taken as features. Two other measurements for calculating the duration were also explored (Vandermeulen et al., 2016). The energy envelope is divided into ten uniformly distributed energy levels. For each energy level the duration is calculated during which the energy of the sound event exceeds that energy level. A line is then fitted through the ten durations obtained and the slope and intercept are calculated. The slope and intercept were also calculated in this study and tested as a feature.

In this study the signal to noise ratio (SNR) of the sound event was explored as feature, however, the SNR is not a feature which specifically defines a cough. The reason it was used is that unclear sound events or events in a noisy environment are often influenced by background noise. Hence, the characteristics can change, which affects the performance of the classification step.

Another feature explored was the energy envelope of the event. As the energy magnitude of an event itself differs due to the SNR and the distance between the microphone and the sound source, energy ratios between different moments in time and between different frequency bands are used as a feature. To apply this technique, the power of the signal was compared for four intervals over the duration of the sound event. The four intervals were equal in duration between the beginning and end of the sound event. Four frequency bands were also chosen, with the frequency bands not being distributed equally but divided into low (<2 kHz), low-middle (2–4 kHz), middle-high (4–7 kHz), and high frequency bands (>7 kHz). The limits of the different frequency bands were chosen by trial and error. Eq. (1) shows how the power of a sound event is calculated:

$$P = \frac{1}{K}\sum_{n=1}^{K}|x(n)|^2 \quad (1)$$

where x(n) is the amplitude of the relevant frequency interval of the nth audio sample ranging from audio sample n = 1 to K, the length of the sequence for which the power needs to be calculated.

A lot of information (e.g. pitch, spectral centroid, and spectral spread) can be extracted from the frequency content of a sound event. In order to characterise the frequency content, multiple features can be calculated. Instead of making use of the feature values themselves, statistics on those features are used as they tend to be more robust to noise (Cortopassi, 2006). As described by Cortopassi (2006), features were based on short time Fourier transforms and the aggregate power and spectrum envelope (Fig. 4). The aggregated power and spectrum envelope were calculated by summing all the power values over frequency and time respectively. Features recommended by Cortopassi (2006) considered in this work were the maximum, median (M), 10th percentile (P1), 90th percentile (P2), interpercentile range (IPR), percentile skewness, concentration, lower value, upper value, lower-upper range, and lower-upper skewness, measured in both time envelope and spectrum aggregate using a particular fraction of 0.8. The M, P1, and P2 can be observed in Fig. 4. In addition to these features, the spectral centroid (Eq. (2)), spectral spread (Eq. (3)), spectral entropy (Eq. (5)), and spectral flux (Eq. (6)) were also investigated. These features were part of the MPEG-7 audio description features (Kim, Moreau, & Sikora, 2005). The spectral centroid indicates the centre of 'gravity' of the spectrum and is defined as:

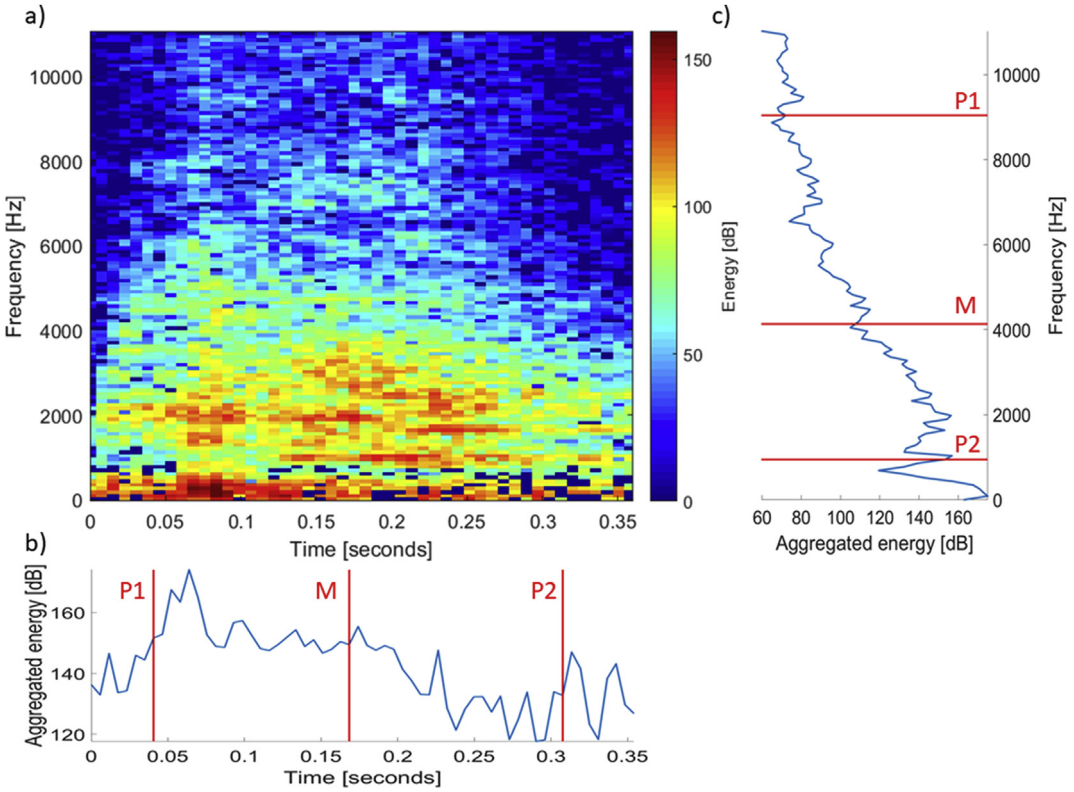


Fig. 4 – From the spectrogram of an example sound (a), the aggregate time envelope (b) and the aggregate frequency spectrum (c) are calculated by summing the short time Fourier transform over frequency and time respectively. On both the aggregated frequency and time spectrum, the median (M), 10th percentile (P1), and 90th percentile (P2) are calculated.

$$C = \frac{\sum_{k=1}^{Wf_L} kX(k)}{\sum_{k=1}^{Wf_L} X(k)} \quad (2)$$

The spectral centroid features were calculated in two different ways. The first approach made use of the aggregate power spectrum where Wf_L was the number of frequency bins used and X(k) was the energy of the kth frequency bin. The second approach looked at the spectral centroid of each short-time spectrum and statistical measures were taken from the time series obtained (mean, median, variance, skewness, and 10-, 20-, 30-, 40-, 60-, 70-, 80-, 90-percentile).

Spectral spread is defined as the average spread of the spectrum in relation to its centroid. In this work the spectral spread was calculated similar as the spectral centroid using the two approaches described above. The spectral spread can be computed using following equation:

$$S = \sqrt{\frac{\sum_{k=1}^{Wf_L} (k-C)^2 X(k)}{\sum_{k=1}^{Wf_L} X(k)}} \quad (3)$$

Spectral entropy indicates how the spectral energy of an audio event is distributed, more specific it is a measure for abrupt changes in the spectral energy. To calculate the spectral entropy, first the normalised spectral energy (Eq. (4)) is calculated by dividing the energy of each frequency bin, k = 1 to Wf_L, through the total spectral energy:

$$n_k = \frac{X(k)}{\sum_{k=1}^{Wf_L} X(k)} \quad (4)$$

Then the spectral entropy can be computed using following equation:

$$H = -\sum_{k=1}^{Wf_L} n_k \log_2(n_k) \quad (5)$$

Spectral flux is a measure of the difference between two successive short-time spectra. It can be computed using following equation where Xi was the ith short-time Fourier transform (Giannakopoulos & Pikrakis, 2014):

$$Fl_{(i,i-1)} = \sum_{k=1}^{Wf_L} (X_i(k) - X_{i-1}(k))^2 \quad (6)$$

2.5.3. *Classification*

Classification is the step where it is decided if an event is automatically identified as a cough event or a non-cough event based on the feature values. As acoustic and noise conditions between calf houses and compartments differ, the algorithm needs to be designed so it can work in different conditions with minimal influence on accuracy. Therefore, this study tried to make use of features which are less

susceptible to noise. A restriction on the SNR was used to remove unclear sounds, as their characteristics are less explicit and therefore induce false classifications. This restriction was manually adjusted by comparing the histograms of the SNR of all coughs with a rating greater than or equal to three with all other sounds. The restriction was set so only outliers from the cough histogram were removed. Too many false classifications could make the algorithm untrustworthy; hence, too many false positives should be avoided. False negatives should also be avoided, however, as more non-cough than cough sounds are expected in commercial environments, the focus is on the false positives.

The classification in this study was performed using the features shown in Table 2. The duration of an event was used to remove events which were too short or too long to be a cough. Based on the results from Ferrari et al. (2010) and the observed coughs in this study, cough from a calf was considered to have a duration longer than 150 ms but shorter than 1.5 s. The thresholds of the other features (Table 2) were adjusted so the targeted performance could be obtained.

2.6. Performance

The performance measurements used to quantify the results of the algorithms consisted of sensitivity (Eq. (7)), precision (Eq. (8)), and specificity (Eq. (9)). It was not possible to determine the specificity of the event selection algorithm or the total algorithm as other events were not labelled, hence the number of true negatives could not be quantified.

$$sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives} \quad (7)$$

$$precision = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ positives} \quad (8)$$

$$specificity = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives\ +\ number\ of\ false\ positives} \quad (9)$$

In order to avoid overfitting the data, 3-fold cross-validation was used to quantify the performance of the algorithm. The data is split into three smaller sets and the algorithm is trained using two sets (66.6% of the data) and validated on the remaining set (33.3%). This is done three times so each set is used as validation set. The average of the three validations was used to measure the performance. More information about cross-validation can be found in the work of Refaeilzadeh, Tang, and Liu (2009).

Table 2 – Features and thresholds used for classifying the sound events, all features are explained in 2.4.2.

Features	
Duration	• Minimum: 150 ms • Maximum: 1.5 s
SNR (to remove unclear events)	• Minimum: 50
Peak frequency	• Maximum: 2200 Hz
Mean frequency	• Maximum: 2400 Hz
10th percentile frequency	• Maximum: 4300 Hz
90th percentile time	• Maximum: 1000 Hz
Lower-upper range frequency	• Minimum: 1500 Hz
Lower-upper range time	• Minimum: 80 ms • Maximum: 310 ms
Frequency density	• Minimum: 2.3
Time density	• Minimum: 0.7 • Maximum: 4.5
Frequency concentration	• Maximum: 11,000 Hz
Interpercentile range time	• Minimum: 50 ms
Maximum of the spectral flux	• Maximum: 140,000,000
Variance of the spectral centroid	• Maximum: 75
Variance of the spectral entropy	• Minimum: 0.0750
90th percentile of the spectral entropy	• Minimum: 3.5 • Maximum: 5.4
Skewness of the spectral spread	• Minimum: −1
40th percentile of the spectral spread	• Minimum: 6
80th percentile of the spectral spread	• Minimum: 7.5 • Maximum: 22
Energy ratios between different frequency bands:	
• 2000–4000 Hz/< 2000 Hz	• Minimum: 0.03
• 2000–4000 Hz/> 7000 Hz	• Minimum: 15
• > 7000 Hz/2000–4000 Hz	• Minimum: 0.0009
• > 7000 Hz/4000–7000 Hz	• Minimum: 0.04
Slope of the durations fitted through the energy levels	• Maximum: 2.75

3. Results and discussion

3.1. Labelling

For this research 445 min of data were analysed and labelled. For every compartment in both batches of calves, 50 min of data were chosen to label. The data were chosen so different moments during the day and calves of different ages were labelled. This is to avoid biasing the algorithm on age or time-dependent background noises. For the first batch, no data were available from compartment one. More sounds (145 min) were labelled for compartment 2 as the initial algorithm was developed on these data. These 145 min were chosen from 8 different days distributed between the beginning and end of the batch. For compartment 3 and 4, five minutes were randomly chosen for each day over a 5 day period. Also, 25 min were chosen from May 4th and May 23rd for compartment 3 and 4, respectively. For the second batch, 25 min of data were labelled for all compartments on May 21st and June 15th spread over the day, more precisely 5 min around 4, 8, 12, 16, and 20.00 h.

In total, 664 coughing events were labelled in the recorded audio files. This gave an average of one cough every 40 s. Table 3 gives a detailed overview of the analysed data and the corresponding total number of labelled calf coughs. For all cases, except compartment four during the first batch (47.6%), more than 50% of the coughs labelled received a label of three or above. In total, 380 coughs (57.2%) received a label greater than or equal to three. Compared to the study of Vandermeulen et al. (2016), where 385 coughs were labelled, the number of coughs with a label of three or above was similar. The studies on pig coughing of Guarino et al. (2008) and Chung et al. (2013)

Table 3 – Overview of the observed data and labelled calf coughs in all compartments for both batches of calves.

Compartment	Batch	Data labelled (minutes)	Labelled calf coughs (all labels)	Labelled calf coughs (label ≥3)
1	1	N/A	N/A	N/A
	2	50	27	18
2	1	145	299	169
	2	50	40	23
3	1	50	149	88
	2	50	52	32
4	1	50	82	39
	2	50	15	11
Total:		**445**	**664**	**380**

used a dataset of 159 coughs and 300 coughs from sick pigs, respectively.

3.2. Event selection

All events in the sound data which were possible cough events were selected. The sensitivity and precision of the event detection algorithm are 99.1% and 3.1% respectively. More detailed results are presented in Table 4, where it can be see that almost all calf coughs are selected. Only during batch 1 for compartments 2 and 3 were 4 and 2 coughs missed, respectively, which is only 1.3% of the coughs during that batch in both compartments. For event selection, precision was considered to be less important as the goal was to detect all possible coughing events without taking into account the classification component.

In total, 20,938 potential cough events were selected by the event selection algorithm with a total time duration of 3431.9 s. The duration of all recordings was 26,700 s, the amount of data that needs to be processed in the next steps is reduced by a factor of 7.78. Only 3.2% of the selected events were cough events. Most cough detection studies only focus on the classification of coughs and work with manual extracted individual sounds often under laboratory conditions. This leads to a higher ratio of cough compared to other sounds: 59.9% (Chedad et al., 2001), 26.9% (Guarino et al., 2008), and 60.0% (Chung et al., 2013). Therefore, our classification needs to have a high specificity to reduce the number of false positives remaining to within acceptable limits.

3.3. Classification

Each individual interval selected by the event selection was classified as either cough or other sound based on the values of the features. The classification is calculated based on the output of the event selection algorithm where 658 cough and 20,280 non-cough events were found. Based on this input data, true negatives could be determined and hence the specificity could be calculated. The sensitivity, precision, and specificity are respectively 41.8%, 94.2%, and 99.9%. If only the coughs with a quality label higher than or equal to three are considered, a sensitivity of 60.4% and a precision of 93.1% are obtained. As coughing occurs far less than other sounds it is important that there is a high specificity to reduce other sounds being falsely classified as coughs. This has a trade-off as more coughs which are unclear would then be rejected. Due to this, the SNR is an important feature to remove events which don't differ much from the background. Also cough events which overlap with other sound events are labelled with a low value (label equals one or two) as the characteristics of the overlapping events are influenced by the other event. Inaccurate reference data, labels equal one or two, do not permit the development of a robust algorithm and therefore should be removed during the development phase.

In Table 5, the detailed results of the algorithm proposed in this paper are presented for each compartment and both batches individually. When only coughs with a label greater than or equal to three are considered, coughs with a label of one or two are ignored and not counted as misclassified. One can observe in Table 5 the low detection ratio in compartment 1 for batch 2. When looking at coughs with a label greater than or equal to three, only 3 coughs out of 18 are correctly classified. Additionally in compartment 4 during batch 2, the classification yielded a precision of only 53.8%. In both cases the labelled data (50 min) is rather limited and further work is needed.

We achieved a lower sensitivity (41.8%) than previously published cough detection algorithms: 85.5% (Guarino et al., 2008), 94.0% (Chung et al., 2013), and 50.3% (Vandermeulen et al., 2016). However, if only clear coughs (label greater than or equal to three) are considered, the sensitivity rises to 60.4%. The precision of the algorithm is in line with the other algorithms proposed in literature (Chung et al., 2013; Guarino et al., 2008; Vandermeulen et al., 2016). The

Table 4 – Results of the event selection algorithm proposed in 2.4.1.

Compartment	Batch	Coughs selected by algorithm vs labelled coughs (all labels)	Error	Coughs selected by algorithm vs labelled coughs (label≥3)	Error	Events selected by algorithm
1	1	N/A	N/A	N/A	N/A	N/A
	2	27/27	0	18/18	0	1908
2	1	295/299	4	168/169	1	7646
	2	40/40	0	23/23	0	2132
3	1	147/149	2	88/88	0	2718
	2	52/52	0	32/32	0	2253
4	1	82/82	0	39/39	0	2212
	2	15/15	0	11/11	0	2069
Total:		**658/664**	**6**	**379/380**	**1**	**20,938**

Table 5 – Results of the classification algorithm proposed in 2.4.3.

Compartment	Batch	True Positives (All labels)	True positives (label ≥ 3)	False positives (All labels)	True negatives (All labels)	False negatives (All labels)	False negatives (label ≥ 3)
1	1	N/A	N/A	N/A	N/A	N/A	N/A
	2	4	3	2	1879	23	15
2	1	149	119	4	7346	146	49
	2	20	15	1	2091	20	8
3	1	56	52	1	2569	91	36
	2	16	16	3	2198	36	16
4	1	23	18	0	2130	59	21
	2	7	6	6	2048	8	5
Total:		**275**	**229**	**17**	**20,261**	**383**	**150**

specificity is higher at 99.9% compared to 86.6% (Guarino et al., 2008), 94.6% (Chung et al., 2013), and 99.2% (Vandermeulen et al., 2016). A higher specificity is needed as there are many more non-coughing sounds in the current experiment. Both the difference in sensitivity and precision between this study and the studies of Guarino et al. (2008) and Chung et al. (2013) can be explained by the way the data was recorded. In both those studies, the recording device was placed close, 20–50 cm (Guarino et al., 2008) and 1 m (Chung et al., 2013), to one animal resulting in high quality sounds.

The classification in this study was based on the selected set of features (Table 2), and for each selected feature a threshold was determined. This threshold was independent of the compartment so the same algorithm could run in the different compartments without the need for fine-tuning on the specific compartment. This is an improvement to the study of Vandermeulen et al. (2016) where a calibration of the reference labels was needed for each compartment. By removing this step, an important difficulty for practical implementation of the algorithms is eliminated. Further research should investigate the influence of different types of calf houses on the performance of the algorithm as the compartments in this study were very similar. Additionally, multiple other parameters should be tested as the acoustic characteristics of the compartments filled with calves depends on many factors including weather conditions, bedding material, number of animals, breed, and age of animals.

3.4. *Performance of the total algorithm*

The performance of the event selection and classification into cough and non-cough sounds has already been given in 3.2 and 3.3, respectively. Here the performance of all steps combined is considered. The inputs are sound recordings in WAV format, thereafter the noise is suppressed, events are selected, feature values are calculated, and the events are classified as either cough or non-cough. The algorithm proposed in this study yields a sensitivity of 41.4% and a precision of 94.2%. If only the coughs with a label equal to or greater than three are considered, the sensitivity is 60.3% and the precision is 93.1%. The sensitivity is similar as the sensitivity for the classification step and the precision is identical. The only difference between the performance of the classification and the total algorithm is that the cough events not selected in the event selection step are counted as false negatives. As there were only 6 coughs out of the 664 coughs not selected, the impact on the sensitivity is small.

For the coughs with a label greater than or equal to three, the total algorithm yields a sensitivity of 16.7%, 70.4%, 65.2%, 59.1%, 50.0%, 46.2%, and 54.5% in the individual compartments over both batches. It could be argued that in most cases around 60% of calf coughs are undetected. This is mainly due to the removal of unclear or less clear coughs and the rather strict classification to limit the number of false positives. It is important to observe that coughs were detected in all compartments for both batches. Therefore the first case (16.7%) should be investigated further to see if and where problems might have occurred. The precision of the individual cases indicate that the algorithm works very well in 5 out of 7 cases with a precision of 97.3%, 95.2%, 98.2%, 84.2%, and 100%. For these experiments the obtained results are good, indicating that the algorithm works in different conditions without an individual calibration step for each compartment and each batch. The precision of compartments 1 and 4 during the second rearing period was lower, 66.6% and 53.8%, respectively, but the precision of 66.6% is still in line with the results of Guarino et al. (2008) (70.1%). More false positives were detected in these compartments during the second batch, and further analysis is needed to determine whether this is a problem of the limited amount of labelled data (50 min for each case). A possible explanation could be that the environmental noise was higher in comparison to the other cases resulting in a lower detection rate. Another possible explanation could be that the specific coughs come from a calf already affected by BRD, changing the characteristics of the cough sound in a similar manner as observed with pigs (Exadaktylos et al., 2008). However, as the objective is to detect respiratory disease, a clear indication about the relative amount of coughing (rising or decaying) should be detected. In both cases there were still more coughs than non-coughs detected, indicating that this maybe is not the problem. The main challenge in each algorithm remains to reduce the number of false positives to minimise the risk that the user will stop using it due to false alarms.

The algorithm proposed in this study is independent of the compartment. The same algorithm could detect coughs in each compartment. Further research in multiple calf housing is needed to verify whether a single algorithm, this one or a similar approach which defines a set of features and determines a fixed threshold, is suitable, independent of the specific house.

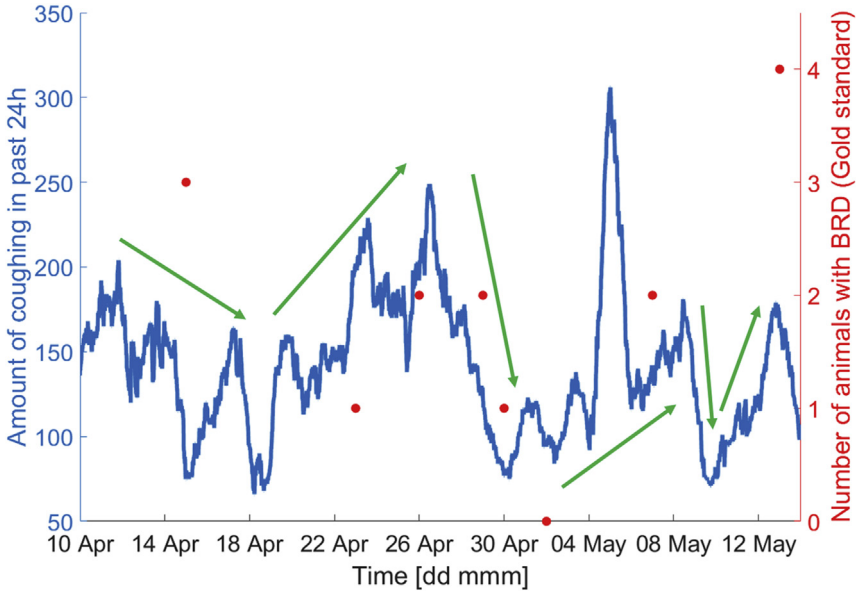


Fig. 5 – Total calf coughs in the past 24 h detected in compartment 4 during batch 1. The points represent the number of calves which presented with BRD according to the gold standard.

3.5. *Comparison between detected coughs and gold standard*

Figure 5 shows the number of calf coughs detected by the algorithm over the time period compared with the gold standard measurements of BRD made in compartment 4. The dots in the figure represent the number of animals diagnosed with BRD during the inspection rounds. It can be seen that between the 18th and 27th of April the trend in the number of calf coughs detected increases, and this relates with the number of calves diagnosed with BRD (one calf on April 23rd and two calves on April 26th). From April 27th the number of detected calf coughs drop and the number of BRD infected calves also decreases (to 0), while between May 4th–7th, the coughing trend rises as two calves were diagnosed with BRD. However, on May 5th an unexplained intense coughing peak can be seen. As only 2 calves were diagnosed with BRD, this coughing bout may have been caused by other reasons, e.g. poor air quality. However, this cannot be confirmed within our dataset.

In general, Fig. 5 shows a relation between the number of calves diagnosed with BRD and the detected coughs. Further research is needed to explore the scope to raise alarms automatically. In addition, there is still quite some variance in the coughs detected, independent of the trend. A next step will be to scrutinise the detected coughs and validate the reason for this variance: possible explanations could be that one type of sound event is being falsely classified or the misclassification of cough events due to environmental noise.

4. Conclusions

This paper proposes an algorithm for the detection of calf coughing in a real-life situation. A quality rating was added to the reference dataset to mitigate the impact of unclear and overlapping cough events. The algorithm yielded a sensitivity of 41.4% and a precision of 94.2% without the need to change parameters for each compartment. Hence, the algorithm is suited to run in different calf compartments. The robustness of the algorithm was evaluated by looking at the performance of the individual compartments during the different time periods. In every compartment during both batches of animals, coughs could be detected, hence the proposed algorithm appears suited to detect coughs in compartments under different circumstances. The precision of the individual experiments indicates that the algorithm works well in 5 out of 7 cases, with a precision ranging from 84.2% to 100%. In the other two cases, the algorithm generates more false positives and further work is required to resolve this. The comparison between the detected coughs and BRD detection using gold standard shows that the trend of the detected coughs increases and decreases when the number of animals diagnosed with BRD increases and decreases, respectively.

Acknowledgements

The authors gratefully acknowledge the European Community for financial support in Collaborative Project EU-PLF, grant number: 311825, under funding scheme FP7-KBBE.2012.1.1-02.

REFERENCES

Aerts, J.-M., Jans, P., Halloy, D., Gustin, P., & Berckmans, D. (2005). Labeling of cough data from pigs for on-line disease monitoring by sound analysis. *Transactions of the American Society of Agricultural Engineers*, 48(1), 351–354.

Allan, E. M., Pirie, H. M., Selman, I. E., & Snodgrass, D. R. (1978). Some characteristics of a natural infection by parainfluenza-3 virus in a group of calves. *Research in Veterinary Science, 24*(3), 339–346.

Amrulloh, Y. A., Abeyratne, U. R., Swarnkar, V., Triasih, R., & Setyati, A. (2015). Automatic cough segmentation from non-contact sound recordings in pediatric wards. *Biomedical Signal Processing and Control, 21*, 126–136.

Aydin, A., Bahr, C., & Berckmans, D. (2015). A real-time monitoring tool to automatically measure the feed intakes of multiple broiler chickens by sound analysis. *Computers and Electronics in Agriculture, 114*, 1–6.

Aydin, A., Bahr, C., Viazzi, S., Exadaktylos, V., Buyse, J., & Berckmans, D. (2014). A novel method to automatically measure the feed intake of broiler chickens by sound technology. *Computers and Electronics in Agriculture, 101*, 17–23.

Barry, S. J., Dane, A. D., Morice, A. H., & Walmsley, A. D. (2006). The automatic recognition and counting of cough. *Cough, 2*(8). https://doi.org/10.1186/1745-9974-2-8.

Berckmans, D. (2004). Automatic on-line monitoring of animals by precision livestock farming. In *Proceedings of the international society for animal hygiene congress – animal production in Europe: The way forward in a changing world. Saint-Malo (France), 11–13 October 2004* (pp. 27–30).

Berckmans, D., Hemeryck, M., Berckmans, D., Vranken, E., & van Waterschoot, T. (2015). Animal sound…Talks! real-time sound analysis for health monitoring in livestock. In *Proceedings of the international symposium on animal environment and welfare. Chongqing (China), October 23–26* (pp. 215–222).

Chedad, A., Moshou, D., Aerts, J.-M., Van Hirtum, A., Ramon, H., & Berckmans, D. (2001). Recognition system for pig cough based on probabilistic neural networks. *Journal of Agricultural Engineering Research, 79*(4), 449–457.

Chung, Y., Oh, S., Lee, J., Park, D., Chang, H.-H., & Kim, S. (2013). Automatic detection and recognition of pig wasting diseases using sound data in audio surveillance systems. *Sensors, 13*, 12929–12942. https://doi.org/10.3390/s131012929.

Cortopassi, K. A. (2006). *Automated and robust measurement of signal features*. The Cornell Lab of Ornithology Bioacoustics Research Program. http://www.birds.cornell.edu/brp/research/algorithm/autoated-and-robust-measurement-of-signal-features. (Accessed 20 November 2015).

Edwards, T. A. (2010). Control methods for bovine respiratory disease for feedlot cattle. *Veterinary Clinics of North America Food Animal Practice, 26*(2), 273–284.

Exadaktylos, V., Silva, M., Aerts, J.-M., Taylor, C., & Berckmans, D. (2008). Real time recognition of sick pig cough sounds. *Computers and Electronics in Agriculture, 63*(2), 207–214.

Fagerlund, S. (2007). Bird species recognition using support vector machines. *EURASIP Journal of Applied Signal Processing, 2007*(1), 64–71. https://doi.org/10.1155/2007/38637.

Ferrari, S., Piccinini, R., Silva, M., Exadaktylos, V., Berckmans, D., & Guarino, M. (2010). Cough sound description in relation to respiratory diseases in dairy calves. *Preventive Veterinary Medicine, 96*(3–4), 276–280. https://doi.org/10.1016/j.prevetmed.2010.06.013.

Giannakopoulos, T., & Pikrakis, A. (2014). *Introduction to audio analysis: A MATLAB approach*. Kidlington (Oxford): Academic Press. ISBN: 978-0-08-099388-1.

Guarino, M., Jans, P., Costa, A., Aerts, J.-M., & Berckmans, D. (2008). Field test of algorithm for automatic cough detection in pig houses. *Computers and Electronics in Agriculture, 62*, 22–28.

Healy, A. M., Monaghan, M. L., Basset, H. F., Gunn, H. M., Markey, B. K., & Collins, J. D. (1993). Morbidity and mortality in a large Irish feedlot; microbiological and serological findings in cattle with acute respiratory disease. *British Veterinary Journal, 149*(6), 549–560.

Johnston, D., Earley, B., Cormican, P., Murray, G., Kenny, D. A., Waters, S. M., et al. (2017). Illumina MiSeq 16S amplicon sequence analysis of bovine respiratory disease associated bacteria in lung and mediastinal lymph node tissue. *BMC Veterinary Research, 13*(118). https://doi.org/10.1186/s12917-017-1035-2.

Jones, M. L., & Allison, R. W. (2007). Evaluation of the ruminant complete blood cell count. *Veterinary Clinics of North America: Food Animal Practice, 23*(3), 377–402.

Kashiha, M., Bahr, C., Ott, S., Moons, C. P., Niewold, T. A., Ödberg, F. O., et al. (2013). Automatic identification of marked pigs in a pen using image pattern recognition. *Computers and Electronics in Agriculture, 93*, 111–120.

Kim, H.-G., Moreau, N., & Sikora, T. (2005). *MPEG-7 audio and Beyond: Audio content indexing and retrieval*. Wiley & Sons.

Martin, R. (2001). Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing, 9*(5), 504–512.

Oczak, M., Ismayilova, G., Costa, A., Viazzi, S., Sonoda, L. T., Fels, M., et al. (2013). Analysis of aggressive behaviours of pigs by automatic video recordings. *Computers and Electronics in Agriculture, 99*, 209–217.

Ozkanlar, Y., Aktas, M. S., Kaynar, O., Ozkanlar, S., Kirecci, E., & Yildiz, L. (2012). Bovine respiratory disease in naturally infected calves: Clinical signs, blood gases and cytokine response. *Revue de Medicine Veterinaire, 163*(3), 123–130.

Pardon, B., Hostens, M., Duchateau, L., Dewulf, J., De Bleecker, K., & Deprez, P. (2013). Impact of respiratory disease, diarrhea, otitis and arthritis on mortality and carcass traits in white veal calves. *BMC Veterinary Research, 9*, 79. https://doi.org/10.1186/1746-6148-9-79.

Potter, T., & Aldridge, B. (2010a). Systematic approach to calf pneumonia. *Livestock, 15*(6), 31–34. https://doi.org/10.1111/j.2044-3870.2010.tb00317.x.

Potter, T., & Aldridge, B. (2010b). Systematic approach to calf pneumonia Part 2. *Livestock, 15*(7), 19–24. https://doi.org/10.1111/j.2044-3870.2010.tb00322.x.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. LIU, & M. T. ÖZSU (Eds.), *Encyclopedia of database systems* (pp. 532–538). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-39940-9_565.

Snowder, G. D., Van Vleck, L. D., Cundiff, L. V., Bennet, G. L., Koohmaraie, M., & Dikeman, M. E. (2007). Bovine respiratory disease in feedlot cattle: Phenotypic, environmental, and genetic correlations with growth, carcass, and longissimus muscle palatability traits. *Journal of Animal Science, 85*(8), 1885–1892. https://doi.org/10.1111/j.2044-3870.2010.tb00317.x.

Tennant, T. C., Ives, S. E., Harper, L. B., Renter, D. G., & Lawrence, T. E. (2014). Comparison of tulathromycin and tilmicosin on the prevalence and severity of bovine respiratory disease in feedlot cattle in association with feedlot performance, carcass characteristics, and economic factors. *Journal of Animal Science, 92*(11), 5203–5213. https://doi.org/10.2527/jas.2014-7814.

Tullo, E., Fontana, I., Diana, A., Norton, T., Berckmans, D., & Guarino, M. (2017). Application note: Labelling, a methodology to develop reliable algorithm in PLF. *Computers and Electronics in Agriculture, 142*, 424–428.

Tullo, E., Fontana, I., & Guarino, M. (2013). Precision livestock farming an overview of image and sound labelling. In *Proceedings of precision livestock farming '13. Joint European conference on precision livestock farming. Leuven (Belgium), 10–12 September 2013* (pp. 30–38).

Van Hirtum, A., Aerts, J.–M., Berckmans, D., Moreaux, B., & Gustin, P. (1999). On–line cough recognizer system. *Journal of the Acoustical Society of America, 106*, 2191–2191.

Vandermeulen, J., Bahr, C., Johnston, D., Earley, B., Tullo, E., Fontana, I., et al. (2016). Early recognition of bovine respiratory disease in calves using automated continuous monitoring of cough sounds. *Computers and Electronics in Agriculture, 129*, 15–26.

White, B. J., & Renter, D. G. (2009). Bayesian estimation of the performance of using clinical observations and harvest lung lesions for diagnosing bovine respiratory disease in post-weaned beef calves. *Journal of Veterinary Diagnostic Investigation, 32*, 446–453.