Computers and Electronics in Agriculture xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Ability evaluation of a voice activity detection algorithm in bioacoustics: A case study on poultry calls

Alireza Mahdavian[a,b], Saeid Minaei[a,*], Ce Yang[b], Farshad Almasganj[c], Shaban Rahimi[d], Peter M. Marchetto[b]

[a] Biosystems Engineering Department, Faculty of Agriculture, Tarbiat Modares University, Tehran 14117-13116, Iran
[b] Department of Bioproducts and Biosystems Engineering, University of Minnesota, Saint Paul, MN 55108, United States
[c] Department of Bioelectrics, Faculty of Biomedical Engineering, Amir Kabir University of Technology, 424 Hafez Ave, Tehran 15875-4413, Iran
[d] Department of Poultry Science, Faculty of Agriculture, Tarbiat Modares University, Tehran 14117-13116, Iran

ARTICLE INFO

Keywords:
Bioacoustics
Health monitoring
Respiratory diseases
Audio features
Gallus gallus

ABSTRACT

Poultry is one of the most strategic source of human foods. There have been seen some hopeful signs of bioacoustics application to monitor the health condition of this vital food source. One of the obstacles is that the bird's call is combined with some unvoiced sounds and extracting the calls is not easy, especially when the bird is sick. This research is a report on successful application of some of the features involved in extracting healthy and non-healthy birds' calls from their sound signals. One hundred and twenty birds from two genotypes – Ross and Cobb – were placed in two groups, a control and those challenged with respiratory diseases. They were reared and their sound was recorded daily. The vocal phrases of the recorded audio signals were extracted using the presented algorithm. Results of analysis showed that an increase in age and onset of illness are two factors that cause an error increase. Detection accuracy was calculated at 95% for healthy young birds and 72% for non-healthy birds. A significant part of this error is due to misclassing the calls as non-vocal segments. This meant that 97% of the activities classified as vocal phrases were, in fact, vocal. These results showed that the idea of such an easy-to-implement algorithm could potentially be employed for the coarselevel segmentation of some animal vocalization signals with reliable outputs, which is an essential and primary step in bioacoustics research.

1. Introduction

Different techniques have been introduced to study domesticated animals, but among them, sound-based precision livestock farming (PLF) has significant advantages over other methods such as cameras or accelerometers. Besides the fact that microphones are contactless and relatively cheap, there is no need for a direct line of sight. Large groups of animals can be monitored with a single sensor (Berckmans et al., 2015). Bioacoustics is a branch of signal processing science that has been developed to discover and decode the acoustic signals emitted from biological systems to know more about their condition. Although all of the signals emitted from natural systems can be placed in this area, the majority of research in this field has been conducted on human and animal audio signals. Studies on the human voice have mostly been aimed at speech recognition, speech reconstruction and audio-respiratory disease detection. In recent years, a number of valuable research studies have focused on animal biological signals such as the ones mentioned below.

Using vocalization signals as a behavior-based welfare indicator to monitor broiler activity is not uncommon (Peña Fernández et al., 2015). For example, the coughing sound has been studied to identify respiratory infection in pigs and dairy calves (Ferrari et al., 2008; Silva et al., 2008), the pecking sound of birds was analyzed to study their feeding behavior (Aydin et al., 2015). Similarly, it has been employed in the vocal recognition of mother–offspring in cattle (de la Torre et al., 2016), in bird population size estimation with the help of bioacoustics (Bardeli et al., 2010), in the biological components of the soundscape, in the way the sounds produced by animals interact with each other (Hildebrand and Baumann-Pickering, 2013), in using sound analyzing techniques for modeling the weight broilers (Fontana et al., 2017), and in the identification of vocal patterns in young broilers (Fontana et al., 2016). Exadaktylos et al. (2014) mentioned many interesting applications of animal sounds where a particular part of the signal is investigated. These studies show that valuable information can be gained from the system under study with the help of biological voices. Among the mentioned areas and animals, the present study, which is part of a

* Corresponding author.
E-mail address: minaee@modares.ac.ir (S. Minaei).

https://doi.org/10.1016/j.compag.2019.105100
Received 23 April 2018; Received in revised form 4 November 2019; Accepted 7 November 2019

more extensive research on poultry health monitoring, is focused on obtaining information from the vocalization signal of broilers (Mahdavian, 2017).

The first step in this process is recording the sound signal of the target (the system that is under study) and of course, it should be noted that an accurate and desirable signal (a signal without any noises or sound activities that have a source other than the target) cannot be expected at this step, since none of the animals, their calls, and their environmental sounds are under our control. This issue can make it more challenging to research animal vocalization rather than the human voice; therefore, the extraction of bird calls from ambient noises and non-vocal sounds is an integral part of research in this area. Ventilation systems or electrical devices can cause ambient noise while short-duration sound activities like the pecking of birds could be considered as non-vocal sounds.

Like most birds, poultry calls include the audio phrases issued by the bird at different intervals. These audio phrases are interwoven between the parts of the silence and signal segments, which are not useful such as noise and non-vocal sound activities. Among these, the best way of dealing with continuous harmonic sounds that have sources other than the birds, such as ventilation noises, would be to design a proportionate digital filter.

Designing such filters requires close inspection of the target signal and is not in the scope of this article, although a band-pass filter that covers the frequency range of chicken calls can be beneficial in this regard. Such a filter was employed in this research (Fig. 1).

Although most of the ambient noises could be removed using a band-pass filter, for extracting bird calls from silence and non-vocal sounds, special attention is necessary. Admittedly, the performance of the algorithm, which is developed to pick up birds' vocal frames, would have a direct and undeniable effect on the performance of any other sound processing system in this area. In other words, introducing a reliable algorithm that can extract vocal sounds among other sound activities in gathered signals, can be helpful in developing the animal bioacoustics field. Such an algorithm is called voice activity detection (VAD).

In different studies, different algorithms are introduced for the course level segmentation of an acoustic signal (Bachu et al., 2008; Rybach et al., 2009; Moattar et al., 2010; Bhandari et al., 2014). These algorithms are designed based on their target signals and their application. The algorithms have different computational complexity, but it can be assumed that in most of them, acoustic features such as short-time energy (STE), aero crossing rate (ZCR), signal to noise ratio (SNR), among others, have had a leading role.

Regardless of the structure of an algorithm and features which are used in that, this is very important to know the performance, strengths, and weaknesses of a VAD algorithm accurately in its working area. Manual checking can identify the real condition of sound activity. Actually, the only simple idea or principle behind a VAD algorithm is that the output of a robust VAD algorithm should not be different from what a human recognizes. In other word, an algorithm should extract vocal parts of a sound signal and should not misdiagnosis other parts as vocal. As mentioned, employing vocal signals of domesticated animals like poultry in their behavior analysis and health monitoring has attracted more attention to itself these days. Therefore, algorithms which are able to extract vocal segments from an audio signal could be helpful in this field. The first step in this way is extracting vocal syllables from the raw signal, and more important is to know how much we can trust the algorithm when it introduces a signal segment as a vocal syllable, as well as how many of the vocal syllables are incorrectly rejected by the algorithm.

The present research tried to pave the way for further studies in this field by introducing and evaluating an easy implementation of the VAD algorithm. This algorithm extracted vocal syllables in sound signals gathered from two groups of broilers, healthy and challenged with respiratory disease. Acoustic features, peak prominence, short-time energy (STE) and wiener entropy (WE) were employed to detect target segments. The performance of the algorithm was evaluated using different indexes, which are introduced in following:

2. Materials and methods

Tests were done with 120 birds made up of two genotypes of Ross and Cobb broilers. The birds had been reared in a standard (commercial poultry farming protocol) temperature, light, feed, and each of them was numbered by individual tags in order to aid identification. The birds were divided into two groups; control and those challenged with bronchitis and Newcastle diseases. Infectious bronchitis and Newcastle diseases are two of the most common respiratory diseases in the world. The control group was vaccinated against common regional diseases, but the two others had been challenged with 10× overdose vaccines for a related disease on the 10th day. The groups were separately kept to avoid contagion. To ensure the bird's health condition, they were monitored using a serological blood test three times during the

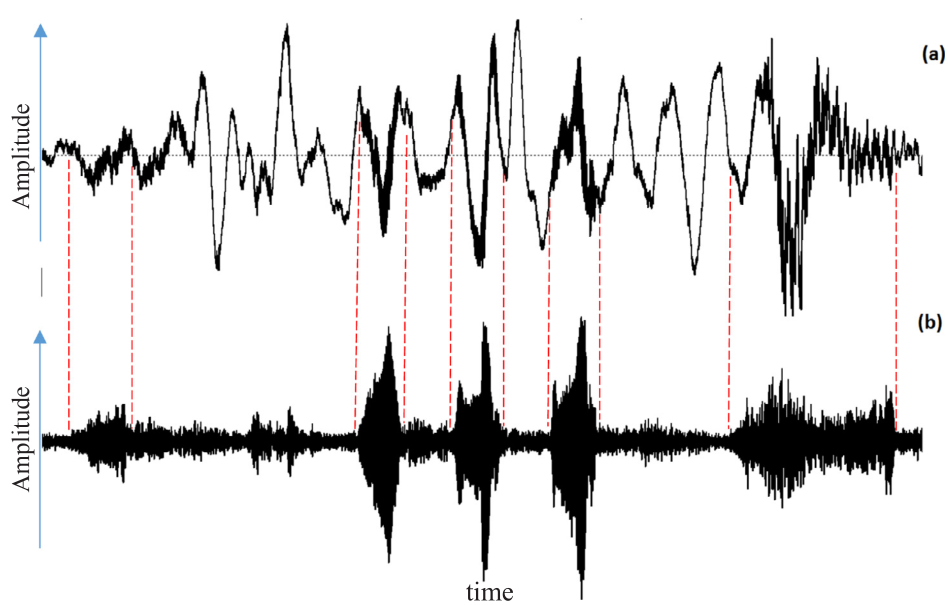

Fig. 1. (a) Raw signal recorded at a poultry house with noise; (b) filtered signal ‘a’ at chickens’ vocal frequency range.

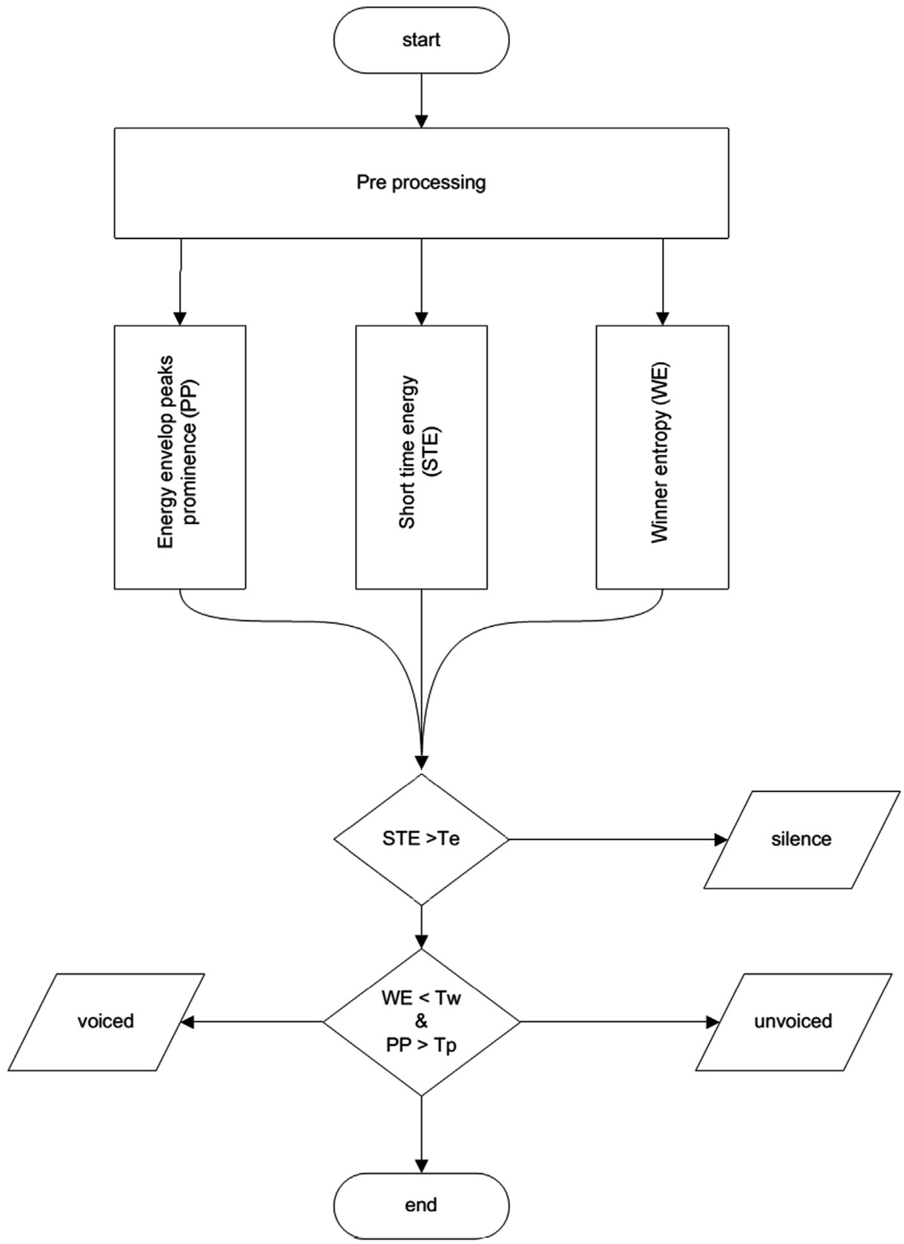

Fig. 2. Flowchart of the presented algorithm; Te, Tw, Tp are threshold values for STE, WE and peak prominence, respectively,which should be adjusted according to the test condition and environmental noises.

program. The sound of each of the birds was recorded daily and individually. It should be mentioned that the disease causes physiological symptoms and that it did not lead to the loss of any of the birds.

The profile of each bird (health condition, genotype, and gender) was traceable using their tag number. Sound signals were recorded separately every day of their growing period for the groups of birds with the same profile. The same protocol was employed in the universities of Tarbiat Modares (Iran) and Minnesota (US) for the experiments and data collection.

The main body of the algorithm employed in this paper is shown in Fig. 2.

T_e in the algorithm is the energy threshold that makes a boundary between the silence and sound activity areas. Actually T_e separates the silent parts of the sound signal and leaves the rest for T_w and T_P which are the thresholds for the energy and peck prominence values, respectively. The algorithm classifies a sound activity in one of two vocal and non-vocal groups based on its WE and CPP values regarding the T_w and T_P thresholds.

2.1. Short-time energy

For a short-term speech signal, an n-th frame window is applied on this signal:

$$x_n(m) = x(m)w(n-m) \quad n-N+1 \le m \le n$$

n = 0, 1 T, 2 T, …, N is the window length, and T is the frame shift.

Short time energy of a signal can be determined from the following expression:

$$E_n = \sum_{m=n-N+1}^{n} [x(m)w(n-m)]^2$$

where w (n-m) is the window, n is the sample that the analysis window is centered on, and N is the window length. The chosen window selects the interim for processing and slides across the progression of squared values. In the presented algorithm, low energy values would be classified as silence and higher values as sound activities.

2.2. Wiener entropy

Wiener entropy or spectral flatness value represents the shape of peaks in a power spectrum and varies between 0 (for spectrum with very sharp peaks) and 1 (for signals with very flat spectrum). A high spectral flatness indicates that the spectrum has a similar amount of power in all spectral bands; therefore, the spectrum envelope would appear relatively flat and smooth.

The Wiener entropy is calculated by dividing the geometric mean of the power spectrum by the arithmetic mean of the power spectrum, i.e.:

$$\mathscr{F}_1 = \frac{(\Pi_{n=0}^{N-1} x(n))^{\frac{1}{N}}}{\frac{1}{N}\sum_{n=0}^{N-1} x(n)} \tag{1}$$

where x(n) is spectrum of the target signal.

2.3. Cepstral peak prominence (CPP)

The CPP measure is the difference in amplitude between the cepstral peak and the corresponding value on the regression line that is directly below the peak (i.e., the predicted magnitude for the frequency at the cepstral peak). The CPP measure represents how far the cepstral peak emerges from the cepstrum background. CPP was successfully employed as a robust and relevant acoustic measure of voice quality in several studies specially when there are some perturbation in the signal because of any problem in vocalization system such (Hartl et al., 2003; Maryn et al., 2009; Kumar et al., 2010; Balasubramanium et al., 2011; Watts and Awan, 2011). For more details on CPP calculations, we refer the readers to Fraile and Godino-Llorente (2014).

With a close look at the flowchart, we can consider the algorithm as a binary classification machine that is looking for silence and non-vocal segments; therefore, parameters listed in Table 1 can be employed to evaluate the performance of the algorithm.

In this table, the true condition is the real nature of the sound activity which was confirmed after manual check labelling procedure; therefore, TPs were unvoiced parts which were correctly identified by the algorithm, FPs were vocal parts which were incorrectly identified as unvoiced, FNs were unvoiced parts which were incorrectly identified as vocal parts, and TNs were vocal parts which were correctly identified by the algorithm.

Using these parameters, we defined the algorithm assessment indicators as bellow:

$$False\ \ posetive\ \ rate\ \ (FPR) = \frac{FP}{\Sigma\ \ true\ \ \text{voiced}\ \ \text{calls}}$$

$$True\ \ negative\ \ rate\ \ (TNR) = \frac{TN}{\Sigma\ \ \text{true}\ \ \text{voiced}\ \ \text{calls}}$$

$$False\ \ omission\ \ rate\ \ (FOR) = \frac{FN}{\Sigma\ \ \text{voice}\ \ \text{identified}}$$

$$Negative\ \ predictive\ \ value\ \ (NPV) = \frac{TP}{\Sigma\ \ \text{voice}\ \ \text{identified}}$$

Table 1
Parameters used as performance criteria of the algorithm.

Total sound phrases		Algorithm's result	
		Unvoiced	Voiced
True condition	Unvoiced sound	True positive (TP) Correctly rejected	False negative (FN) Incorrectly accepted
	Voiced call	False positive (FP) Incorrectly rejected	True negative (TN) Correctly accepted

Table 2
Results of ANOVA indicating the effect of gender, genotype and health condition of chickens on classification accuracy (ACC).

Source	Degree of freedom	Type III sum of squares	Mean square
Corrected model	7	2387a	341
Intercept	1	172,042	172,042
Gender	1	0.66	0.66^{ns}
Genotype	1	150	150^{ns}
Health condition	1	1908	1908**
Gender × Genotype	1	96	96^{ns}
Gender × Health	1	8.1	8.1^{ns}
Genotype × Health	1	20.1	20.1^{ns}
Gender × Genotype × Health	1	204	204^{ns}
Error	16	954	59
Total	24	175,384	
Corrected total	23	3341	

** : Significant on the 1% probability level; ns: none significant.

$$Accuracy\ \ (ACC) = \frac{Algorithm\ \ true\ \ resuls}{Total\ \ sound\ \ phrases}$$

3. Results and discussion

In order to investigate the effect of gender, genotype and health condition of the chicken on the results, the variation of algorithm's ACC in different treatments was analyzed, and the results are presented in Table 2. The experiment design was factorial based on a randomized block (RBD). The treatment focus points were gender (two levels, female and male), genotype (two levels, Ross and Cobb) and health condition (two levels, healthy and non-healthy).

According to Table 2, health condition is the only factor affecting ACC in the study area of this project.

The Broilers' vocal signal changed with the physiological changes in the birds. Therefore for a more accurate assessment, the indicators were studied in three periods of the bird life individually. These periods included age 1–10, 11–20 and 21–42. Faster changes in the audio features are the reason for a broader time resolution in the first half of the bird's growing period.

For evaluating the algorithm, 50 sound segments - each of them with a 3-minute signal - contained several sound phrases that were used as the experiment's input signals, and the outputs were evaluated in a human-made way.

Accuracy is the most general index in this assessment. This is the summation of true acceptations and rejections divided into the number of all of the evaluated phrases. As can be seen in Fig. 3, with the bird's age increasing as well as shifting the health condition from healthy to sick, the accuracy decreased. The responsible error is oriented from two sources. Error *type I* considers the bird's call as an unvoiced sound, and this is a false rejection. This error is the previously mentioned FPR,

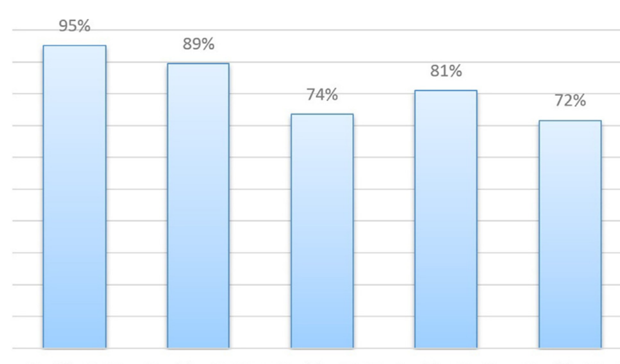

Fig. 3. Accuracy of the algorithm in extracting birds' calls.

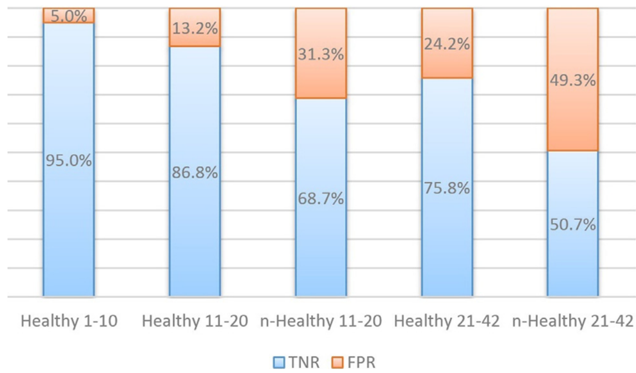

Fig. 4. Specificity showing correctly identified calls (TNR) vs. incorrectly rejected ones (FPR).

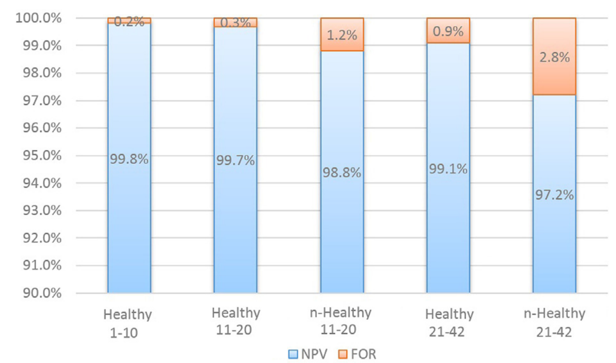

Fig. 5. Showing the correctly accepted voiced calls (NPV) vs. incorrectly accepted ones (FOR).

which caused a decrease in the specificity of the algorithm, as shown in Diagram 3. The error increasing trend in this diagram is compatible with ACC, so almost 50% of the calls were considered as unvoiced and incorrectly eliminated by the algorithm in the second half of the sick bird's age range. These statistics reflect the algorithm's strictness in terms of accepting a sound activity as a voice.

Based on the observations during the research, the bird's voice activity and the call phrase energy would decrease according to the increase in age. Especially in the second half of the bird's growing period, we can hear many chopped calls with a short length and low energy content. Most of them cannot pass through the first filter as it has a short time energy threshold.

We can consider the variations in call frequency range due to both the age increase and due to catching a respiratory disease to be the second reason for the misclassification. Both age and sickness cause more distortion in the bird calls, and this means that the call would be more similar to the unvoiced sounds in terms of Wiener entropy. Some of the calls pass the STE operator and do not pass through the entropy comparator. Thus, they are considered as unvoiced sound incorrectly.

The second might be an accrued error (error *type II*), where it considers unvoiced sounds as a call. This error can decrease the precision of the system and its values during the growing period. This can be seen for both healthy and non-healthy birds as in Fig. 5.

In the case of accruing error *type I*, we lost some of our samples, which was bird calls, but in case of *type II,* the algorithm would feed the wrong input data to the next chain link, which might be a health monitoring system. Although this event is considered as a fault in the system, but this error does not cause any severe problem and more recording is enough to compensate. For error *type II*, due to the algorithm's final application, this error should be considered as the primary error. The trend of this error is also very similar to error *type I,* and we can explain it in the same way. Figs. 3–5 show that although with aging and illness onset the algorithm would lose some of the true calls. This classification strictness leads to a remaining FOR error of less than 3% which means that even in an acute condition, more than 97% of the voice activities selected by the algorithm were bird calls and this can be a reliable result.

4. Conclusion

The health condition monitoring of poultry has always been considered an essential subject for several reasons such as the importance of chicken as an important human food supply and the disease outbreak rate. Therefore any tools introduced to the farmers and veterinarians to help them in this field would be valuable. At the first step of developing a bio-acoustical tool, bird calls should be separated from other unvoiced sounds and silent parts. To achieve this, there are several applicable sound features, some of which are more complicated than others resulting in considerable computational load. This research was undertaken to implement an algorithm with the help of short time energy and wiener entropy features. Since the algorithm can be used in a health monitoring system, performance evaluation of the system was implemented by sampling both healthy and respiratory-disease challenged birds calls. The results are satisfying, and the impurity of the output signals does not exceed 3% even in the condition where bird calls were distorted due to sickness and age. Nevertheless, losing 49% of the calls in the mentioned condition was the cost of increasing signal purity. Hence, more research activity to improve the algorithm and to decrease this error would be valuable. Furthermore, as mentioned in the introduction section, another step to make the research more complete and operational is to design an appropriate filter to make such an algorithm more robust against different types of environmental noise. This would be feasible through an accurate understanding of the sound features of chicken calls.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Tarbiat Modares University and University of Minnesota Faculty Start Up fund for which we are grateful. The authors would like to express their thanks to *Dr. Sally L. Noll and Dr. Anup Kollanoor Johny* for sharing their knowledge during this research.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2019.105100.

References

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. Comput Electron Agric 114, 1–6.
Bachu, R.G., Kopparthi, S., Adapa, B., Barkana, B.D., 2008. Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal. Am. Soc. Eng. Educ. Zo. Conf. Proc. 1–7.
Balasubramanium, R.K., Bhat, J.S., Fahim, S., Raju, R., 2011. Cepstral analysis of voice in unilateral adductor vocal fold palsy. J. Voice 25, 326–329.
Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.-H., Frommolt, K.-H., 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recognit. Lett. 31, 1524–1534.
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: Proc. Int. Symp. Anim. Environ. Welfare. Chongqing, China, Oct. pp. 215–222.
Bhandari, G.M., Kawitkar, R.S., Borawake, M.P., 2014. Audio segmentation for speech

recognition using segment features. In: ICT Crit. Infrastruct. Proc. 48th Annu. Conv. Comput. Soc. India-Vol II. Springer, pp. 209–217.

Exadaktylos, V., Silva, M., Berckmans, D., 2014. Automatic identification and interpretation of animal sounds, application to livestock production optimisation. Soundscape Semiot. Categ.

Ferrari, S., Silva, M., Guarino, M., Aerts, J.M., Berckmans, D., 2008. Cough sound analysis to identify respiratory infection in pigs. Comput. Electron. Agric. 64, 318–325.

Fontana, I., Tullo, E., Carpentier, L., Berckmans, D., Butterworth, A., Vranken, E., Norton, T., Berckmans, D., Guarino, M., 2017. Sound analysis to model weight of broiler chickens. Poult. Sci. 96, 3938–3943.

Fontana, I., Tullo, E., Scrase, A., Butterworth, A., 2016. Vocalisation sound pattern identification in young broiler chickens. Animal 10, 1567–1574.

Fraile, R., Godino-Llorente, J.I., 2014. Cepstral peak prominence: A comprehensive analysis. Biomed. Signal Process Control 14, 42–54.

Hartl, D.M., Hans, S., Vaissière, J., Brasnu, D.F., 2003. Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia. Eur. Arch. Oto-rhino-laryngology 260, 175–182.

Hildebrand, J., Baumann-Pickering, S., 2013. Bioacoustic contributions to the soundscape. J. Acoust. Soc. Am. 134, 4146.

Kumar, B.R., Bhat, J.S., Prasad, N., 2010. Cepstral analysis of voice in persons with vocal nodules. J. Voice 24, 651–653.

de la Torre, M.P., Briefer, E.F., Ochocki, B.M., McElligott, A.G., Reader, T., 2016. Mother–offspring recognition via contact calls in cattle, Bos taurus. Anim. Behav. 114, 147–154.

Mahdavian, A., 2017. Acoustic pattern recognition of broilers challenged with infectious bronchitis and Newcastle viruses. Dissertation, Tarbiat Modares University, University of Minnesota.

Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., Corthals, P., 2009. Acoustic measurement of overall voice quality: A meta-analysis a. J. Acoust. Soc. Am. 126, 2619–2634.

Moattar, M.H., Homayounpour, M.M., Kalantari, N.K., 2010. A new approach for robust realtime voice activity detection using spectral pattern. Acoust. Speech Signal Process. (ICASSP), 2010 IEEE Int. Conf. IEEE, pp 4478–4481.

Peña Fernández, A., Tullo, E., Exadaktylos, V., Vranken, E., Guarino, M., Berckmans, D., 2015. Broiler activity and distribution as behavior-based welfare indicators. Precis. Livest. Farming'15 208–217.

Rybach, D., Gollan, C., Schluter, R., Ney, H., 2009. Audio segmentation for speech recognition using segment features. Acoust. Speech Signal Process. 2009. ICASSP 2009. IEEE Int. Conf. IEEE, pp. 4197–4200.

Silva, M., Ferrari, S., Costa, A., Aerts, J.M., Guarino, M., Berckmans, D., 2008. Cough localization for the detection of respiratory diseases in pig houses. Comput. Electron. Agric. 64, 286–292.

Watts, C.R., Awan, S.N., 2011. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. J. Speech, Lang Hear Res. 54, 1525–1537.